Doctoral Thesis

Magnetic Tweezers

Applications to the study of DNA-protein interactions in DNA replication, condensation, and segregation

Submitted by

César López Pastrana

In partial fulfillment of the requirements for the degree of
Doctor in Biophysics

Department of Condensed Matter Physics

Universidad Autónoma de Madrid

October 2017

A mis padres

Abstract

Magnetic tweezers (MT) is a single molecule technique capable of applying forces and torque to individual DNA molecules. Its simple design and its ability to obtain measurements from different molecules in parallel have positioned MT among the most powerful force spectroscopy techniques. The applications of MT span from DNA mechanics and topology to the study of biomolecular interactions. This dissertation presents a complete description of MT, with emphasis in the methods and procedures for its application to study DNA-protein interactions. We have used MT to investigate two different biological systems in the context of the organization and replication of the bacterial genome.

In a first project, a novel condensation activity of the *B. subtilis* protein ParB is studied. ParB proteins bind *parS* sequences and recruit accessory proteins to coordinate bacterial chromosome segregation during cell division. Surprisingly, we find a highly non-specific DNA binding mode independent on the *parS* sequence and leading to non-organized DNA-protein complexes. The non-specific mode of binding results in DNA condensation at permissive forces below 2 pN. Our results indicate that condensation is the result of bridging between distant DNA regions, as reported by the stabilization of juxtaposed DNA molecules *in cis* and *in trans*.

We studied the implications of the ParB C-terminal domain (CTD) dimerization in DNA binding and condensation, finding a novel DNA binding site at the aforementioned domain. Interestingly, DNA condensation is impaired when ParB is co-incubated with the CTD. The inhibitory activity of ParB condensation by the CTD remains unaltered by mutations at the non-specific binding site of the CTD nor under ionic conditions preventing non-specific binding. Therefore, these results evidence the dependency of CTD dimerization for the condensation activity of ParB.

In a second project, we studied the topological and mechanical determinants of RepC nicking activity. The protein RepC is the initiator of the rolling-circle replication in the pT181 plasmid family. RepC nicks the DNA near the origin of replication at regions presumably forming cruciform structures. We analyzed the nicking and religation activities of RepC in torsionally-constrained molecules. We found that RepC nicks negatively supercoiled DNA and presents an inefficient religation activity after nicking. Remarkably, the nicking activity is force and twist dependent. We characterized the nicking activity as a result of the supercoiling degree and stretching force, finding results consistent with the formation of a cruciform structure. Nonetheless, contrary to previously proposed models, our data suggest a passive role of RepC in the extrusion of the cruciform. Based on homology modelling and our single-molecule experiments we propose a model for the RepC nicking activity dependent on the extrusion of a cruciform as a result of DNA supercoiling.

The results presented in this thesis help to understand the organization of the bacterial chromosome and the topological implications for the initiation of the replication in bacterial plasmids.

Resumen

Las Pinzas Magnéticas (PM) son una técnica de molécula única con capacidad para aplicar fuerzas y torque a moléculas individuales de ADN. Su diseño simple y su capacidad para obtener medidas en paralelo las ha posicionado entre las técnicas más potentes de espectroscopía de fuerzas. Las aplicaciones de las PM abarcan desde mecánica y topología del ADN al estudio de interacciones biomoleculares. Esta disertación presenta una descripción completa de las PM, con un énfasis en los métodos y los procedimientos para su aplicación al estudio de las interacciones entre ADN y proteínas. Hemos usado PM para investigar dos sistemas biológicos en el contexto de la organización y la replicación del genoma bacteriano.

En un primer proyecto, se estudió una novedosa actividad condensadora de la proteína ParB de *B. subtilis*. Las proteínas ParB se unen a las secuencias *parS* y reclutan proteínas accesorias para coordinar la segregación del cromosoma durante la división celular. Sorprendentemente, encontramos un modo de unión altamente no específico independiente de la secuencia *parS* dando lugar a complejos ADN-proteína no ordenados. El modo de unión no específico desencadena en condensación del ADN para fuerzas menores a 2 pN. Nuestros resultados indican que la condensación es el resultado del anclaje entre regiones distantes del ADN debido a interacciones proteína-proteína, indicado por la estabilización de moléculas de ADN yuxtapuestas en *cis* y en *trans*.

También estudiamos las implicaciones del dominio de dimerización C-terminal (DCT) en las actividades de unión a ADN y condensación, y encontramos un nuevo dominio de unión a ADN en dicho dominio. Es interesante notar que la condensación de ParB se inhibe en presencia del DCT. La actividad inhibitoria del DCT sobre la actividad de ParB no se ve alterada por mutaciones en el sitio de unión no específico del DCT ni por condiciones iónicas que previenen este tipo de unión. Por tanto, estos resultados evidencian la dependencia de la dimerización para la actividad condensadora de ParB.

En un segundo proyecto, estudiamos los determinantes topológicos y mecánicos de la actividad de corte de RepC. La proteína RepC es responsable de la iniciación de la replicación por círculo rodante en la familia de plásmidos pT181. RepC corta en hebra simple en una región del ADN cerca del origen de replicación, región que presumiblemente forma estructuras cruciformes. Analizamos las actividades de corte y de religación de RepC en moléculas torsionalmente constreñidas. Encontramos que RepC corta ADN negativamente superenrollado y presenta una baja actividad de religación del corte. Además la actividad del corte es dependiente del grado de superenrollamiento y de la torsión. Caracterizamos dicha actividad en función del grado de superenrollamiento y de la fuerza, encontrando resultados consistentes con la formación de una estructura cruciforme. No obstante, al contrario que en otros modelos propuestos, nuestros datos sugieren un papel pasivo de RepC en la extrusión de la estructura cruciforme. Basándonos en nuestros datos de PM y en modelos de homología, proponemos un modelo para la actividad de corte de RepC debido a la extrusión de una estructura cruciforme como resultado del superenrollamiento del ADN.

Los resultados presentados en esta tesis ayudan a comprender la organización del cromosoma bacteriano y las implicaciones topológicas en la iniciación de la replicación de plásmidos bacterianos.

Contents

I Essential Concepts 5

1 Introduction 7
- 1.1 About biophysics . . . 7
- 1.2 Molecular biophysics . . . 8
- 1.3 The single-molecule perspective . . . 9
  - 1.3.1 Single-molecule fluorescence . . . 9
  - 1.3.2 Atomic Force Microscopy . . . 10
  - 1.3.3 Optical Tweezers . . . 12
  - 1.3.4 Magnetic Tweezers . . . 13
  - 1.3.5 Tethered Particle Motion . . . 14
  - 1.3.6 Trends in single-molecule biophysics . . . 14
- 1.4 Protein-DNA interactions . . . 15
- 1.5 Overview of this thesis . . . 15
- References . . . 17

2 Nucleic acids 21
- 2.1 Introduction . . . 21
- 2.2 DNA Structure . . . 22
  - 2.2.1 Primary structure . . . 22
  - 2.2.2 Secondary and tertiary structures . . . 23
- 2.3 Quaternary structure: DNA Supercoiling . . . 25
- 2.4 Polymer models . . . 28
  - 2.4.1 Kratky-Porod and Freely-jointed chain models . . . 28
  - 2.4.2 Worm-like chain model . . . 29
- 2.5 Nucleic acid-protein interactions . . . 30
  - 2.5.1 Non-specific interactions . . . 31
  - 2.5.2 Specific interactions . . . 31
- References . . . 34

3 Magnetic tweezers 37
- 3.1 Introduction . . . 37
- 3.2 Experimental set up . . . 38
  - 3.2.1 Magnetic Tweezers setup used in this thesis . . . 40
- 3.3 The DNA-bead system and force calibration . . . 41
- 3.4 Determination of bead's coordinates . . . 42
  - 3.4.1 Analysis of in-plane positions . . . 43
  - 3.4.2 Analysis of axial positions . . . 43
- 3.5 Calculation of forces and bead's equation of motion . . . 46
  - 3.5.1 Equation of motion . . . 47
  - 3.5.2 Camera corrections . . . 50
- 3.6 MT experiments under force and torque . . . 52
- 3.7 Magnetic tweezers improvements . . . 55

3.7.1 Enhancement of the applied forces . . . 55
3.7.2 Freely Orbiting Magnetic Tweezers . . . 58
3.7.3 Magnetic Torque Tweezers . . . 60
3.8 Concluding remarks and future directions . . . 62
References . . . 62

II The relation between bacterial chromosome condensation and segregation: the protein ParB as a model 67

4 DNA condensation by the bacterial protein ParB 69
4.1 Introduction . . . 69
4.1.1 DNA condensation . . . 69
4.1.2 Chromosome segregation . . . 71
4.1.3 The ParABS system . . . 71
4.2 Materials and Methods . . . 74
4.2.1 Protein purification . . . 74
4.2.2 Electrophoretic mobility-shift assays (EMSA) . . . 74
4.2.3 DNA substrates . . . 74
4.2.4 Magnetic tweezeers conditions . . . 75
4.3 Results . . . 76
4.3.1 Non-specific DNA binding by ParB leads to condensation . . . 76
4.3.2 ParB does not form ordered structures on the DNA . . . 83
4.3.3 ParB stabilises looping interactions between distal DNA segments . 85
4.4 Discussion . . . 87
References . . . 90

5 The implications of ParB domains in DNA binding and condensation 93
5.1 Introduction . . . 93
5.2 Materials and Methods . . . 95
5.2.1 Protein purification . . . 95
5.2.2 Magnetic Tweezers . . . 95
5.3 Results . . . 96
5.3.1 The NTD is essential for DNA condensation but *parS* recognition is dispensable . . . 96
5.3.2 The CTD binds non-specifically via lysine-rich surface . . . 98
5.3.3 DNA binding by the CTD is essential for DNA condensation and bridging . . . 99
5.3.4 The CTD can both inhibit the formation and decondense ParB-DNA networks . . . 100
5.4 Discussion . . . 101
References . . . 105

III Initiation of the Rolling-circle replication: example of a topology-dependent reaction 109

6 Force and twist dependence of RepC nicking activity 111
6.1 Introduction . . . 111
6.1.1 Rolling-Circle replication of plasmids . . . 111

6.2 Materials and methods . . . 114
6.2.1 Proteins . . . 114
6.2.2 DNA substrates . . . 114
6.2.3 Magnetic tweezers assays . . . 115
6.2.4 Prediction of the atomic structure of RepC dimer . . . 115
6.2.5 Corrections to nicking probablities from static RepC nicking experiments . . . 116
6.3 Results . . . 116
6.3.1 Nicking and religating activities of RepC protein . . . 116
6.3.2 RepC nicking activity is twist-dependent . . . 118
6.3.3 RepC nicking activity is force-dependent . . . 122
6.3.4 RepC-DNA interactions during nicking and religation activities . . . 124
6.3.5 Model for DNA supercoiling-dependent RepC nicking . . . 127
6.4 Discussion . . . 129
References . . . 133

IV Conclusions and future perspectives 139

7 Conclusions and future perspectives 141
7.1 The role of ParB in chromosome condensation . . . 141
7.1.1 Conclusions . . . 141
7.1.2 Future perspectives . . . 142
7.2 Force and twist dependence of RepC nicking activity . . . 143
7.2.1 Conclusions . . . 144
7.2.2 Future perspectives . . . 145

8 Conclusiones y perspectivas futuras 147
8.1 El papel de ParB en la condensación del cromosoma . . . 147
8.1.1 Conclusiones . . . 147
8.1.2 Perspectivas futuras . . . 148
8.2 Dependencia de fuerza y torsion en la actividad de RepC . . . 149
8.2.1 Conclusiones . . . 150
8.2.2 Perspectivas futuras . . . 151

V Appendixes 155

A Force-extension relation in the Freely-jointed chain model 157
A.1 Introduction . . . 157
A.2 Derivation . . . 157
A.3 References . . . 159

B Allan variance analysis of MT spatial and temporal resolution 161
B.1 Introduction . . . 161
B.2 Results . . . 161
B.3 References . . . 163

C A simple model to characterize drag and DNA extension in the presence of flow and magnetic forces 165

C.1 Introduction . . . 165
C.2 Model . . . 165
C.3 References . . . 169

D Implementation of C/C++ libraries to run under LabVIEW 171
D.1 Introduction . . . 171
D.2 Methods . . . 171
C side . . . 172
LabVIEW side . . . 173
D.3 Conclusions . . . 175
D.4 References . . . 177

E List of abbreviations 179

Part I

Essential Concepts

1

Introduction

Chapter overview. What is biophysics? Why to study molecules one-by-one? What is the utility of single molecule techniques in the biological sciences? Every reader of this thesis can be classified between those knowing the answers to these questions and those who aim to answer them. This chapter is focused on the latter and its objective is to give a general and simple perspective of the single-molecule biophysics approach. This will facilitate the understanding of the motivation and the methodologies explained in the following chapters.

1.1 About biophysics

Biophysics is a multidisciplinary science that studies biological systems using the tools and methods from physics. From a historical point of view, there always has been a strong link between physics and biology, with crossed contributions among both sciences. A beautiful example of this connection of both disciplines is embodied in the figure of Robert Hooke (1635-1703), who developed the spring law that carries its name but also coined the term *cell* for describing life units. The first half of the 20th century was especially important for the future development of modern biophysics. For instance, in 1925 Lotka and Volterra independently described the couple of differential equations explaining the dynamics in prey-predator ecology [Lotka, 1920; Volterra, 1926]. Other key event took place in 1944, when Erwin Schrödinger published *What is life?* [Schrödinger, 1993]. In his book, the Nobel laureate disucusses biological questions from a physicist point of view but more importantly, Schrödinger's book inspired many physicists to migrate to biology. Among them was Francis Crick, one of the co-discoverers of the double helix structure of the DNA, who previously had worked in the research of magnetic and acoustic mines during the World War II. In 1952, one year before the determination of the structure of the DNA, Alan Turing –whose contribution in deciphering Nazis' cryptographic codes is well known– applied his mathematical training to biological systems and published a now classical work where the formation of patterns during embryogenesis is described analytically [Turing, 1952].

Already established as a discipline, the contributions of biophysics continued during the second half of the 20th century and in 1958 the Biophysical Society was founded to promote the interaction among scientists of the field all over the world.

During these years there were a vast number of outstanding works in the field of biophysics. To cite some examples of these prolific times, DNA sequencing and the patch-clamp technique (Nobel prizes in 1980 and 1991, respectively) were contributions by biophysicists and their improvements continue up to the present day.

Biophysics at the end of the 20th century was marked by the advent of single-molecule techniques and systems biology. The latter considers the properties of certain entities -from molecules to organisms- including their interaction and analyze them as a whole. The interplay amongst components results in emerging properties which can be essential for the comprehension of complicated and important biological problems [Chuang et al., 2010]. In contrast, single-molecule biophysics, as its name indicates, is focused on the particular physicochemical features of individual molecules, aiming to understand the laws and working mechanisms and their implications on the living organism [Bustamante et al., 2000; Ritort, 2006]. The details of single-molecule techniques will be expanded in section 1.3.

We can just speculate about future directions of biophysics, but probably this will be the century of the so-called *big data*. Improvements of the experimental techniques, the development of new techniques and the rise of the computational power facilitate the production of data. Nonetheless, the generation of enormous quantities of information at both the molecular and the cellular level will lead to new challenges to manage, analyze and contextualize these data. This requires the combination of different scientific approaches including theoretical, computational and experimental point of views.

In summary, biophysics is an interdisciplinary science spanning a wide range of topics and scales where cross-talk between disciplines is a requirement for the generation of high-quality knowledge.

1.2 Molecular biophysics

Molecular biophysics is the branch of biophysics covering the basic building blocks of living systems: proteins, nucleic acids, lipids and carbohydrates, including interactions among them and with other small molecules and ions. There are a wide variety of techniques and tools used in molecular biophysics research. Traditionally, molecular biophysics was mostly represented by structural biology and thus the techniques used were those from that field, mainly electron microscopy, nuclear magnetic resonance, and X-ray diffraction, as well as calorimetry, circular dichroism or infrarred spectroscopy. However, the implementation of molecular dynamics simulations for biomolecules in the 1970s decade resulted in a new field of study by providing atomic insights at a spatial and temporal scale previously inaccessible [McCammon et al., 1977; Karplus and McCammon, 2002]. It was not possible to experimentally deal with the scales of computer simulations until the invention of so-called single-molecule techniques during the 1980s. These techniques have spatial and temporal resolution on the nanometer and the millisecond range, respectively [Neuman and Nagy, 2008] and are critical for the quantitative determination of dynamical, structural and chemical properties of proteins and DNA.

1.3 The single-molecule perspective

In the classical molecular biology, molecular biophysics or chemistry experiments, data are extracted from a large number of molecules. Consecuently, the results represent just an average of the global behaviour of the system of study. This implicit averaging can result in misleading information. A clear example of the perils of averaging occurs in bimodal distributions. In this scenario, the global mean value lies in a non-reppresentative low probability situation between the two populations. The progresses in nanotechnology and instrumentation paved the way for the development of new techniques for manipulating one molecule at a time. These methods allow a direct determination of the distribution of the observables of interest and characterize the heterogeneities of the population. In fact, even in homogeneous populations there are variations from molecule to molecule [Kelley et al., 2001]. Note that biological molecules carry out their function in an environment dominated by thermal fluctuations. The kinetic energy for a molecule at room temperature is on average $3/2k_BT$ ($1\ k_BT$ = 4.1 pN nm). Considering that the energy of an hydrogen bond can be as low as $2k_BT$ [Milo and Philips, 2015], it is obvious that noise is intrinsic to the activity of such machines. Hence, there is a large variability on the physical observables along the time [Kelley et al., 2001] and reaction synchronization is hard to be achieved in ensemble experiments. This makes difficult to study dynamics and derived quantities using volume approaches. Fortunately, single-molecule experiments overcome this problem by capturing the dynamics and the specific properties for each independent entity. The majority of the current existing single-molecule techniques use the force or photons from fluorescent dyes as a probe. Here, some of the main single-molecule techniques are going to be briefly reviewed. These techniques include single-molecule fluorescence, and in addition force spectroscopy techniques as atomic force microscopy (AFM), optical tweezers (OT) or magnetic tweezers (MT) or tether particle motion (TPM).

1.3.1 Single-molecule fluorescence

There are multiple ways to achieve single-molecule resolution in fluorescence microscopy. In the most elementary case, the sample is illuminated based on the principle of total internal reflection (TIR). TIR fluorescence microscopy relies on illuminating the sample with an incident angle higher than the critical angle (Figure 1.1a). This generates an exponentially-decaying evanescent wave that only reaches a few hundreds of nm from the experimental surface and hence the excitation of fluorescence probes is limited to that volume [Axelrod, 1981]. Remarkably, TIR microscopy can be combined with other methods to implement superresolution microscopy or fluorescence resonance energy transfer (FRET). Regarding TIR-based superresolution fluorescent microscopy, the system can require the implementation of additional optics to decrease difraction-limited background fluorescence (e.g. STED) and the combination with mathematical and computational models for image analysis and reconstruction (e.g. PALM/STORM microscopy). In FRET microscopy, a donor fluorophore can be excited under the appropriate wavelenght, and non-radiatively trans-

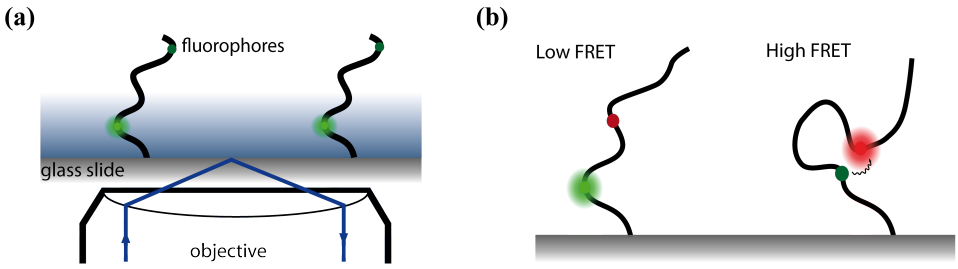

Figure 1.1: Single-molecule fluorescence techniques. a) Objective-based total internal reflection (TIR) microscopy. The sample is illuminated with an angle higher than the critical angle to reach TIR conditions. This leads to the generation of an evanescence wave (blue gradient) which intensisty decays exponentially inside the experimal chamber. Only fluorophores (green dots) located on a region of the molecule (solid lines) inside the evanescence wave are excited emiting photons. b) TIR-based FRET. Donor fluorophores (green dot) are excited with the appropiate wave lenght (left). The presence of aceptor fluorophores (red dots) in the vicinity of the donor results in an energy transfer from the donor to the aceptor. This quenches the donor, exciting the aceptor fluorophore (right).

fer energy to an acceptor fluorophore on its vicinity having an excitation wavelenght different to that of the donor (Figure 1.1b). Since this is only possible at distances $<$ 10 nm (energy transference decays with the distance to the power of six), the FRET-dependent emision of the aceptor can be used as a *spectroscopic ruler* to measure molecular interactions [Roy et al., 2008]. Single-molecule fluorescence techniques are powerful methods that have provided important results. In particular, some beautiful examples are the dermination of the angular step-wise motion of the F1-ATPase [Yasuda et al., 2001] or the behaviour of the DNA at short-scales [Vafabakhsh and Ha, 2013].

Due to the vast number of possibilities, here we have only reviewed a couple of examples of the multiple approximations existing to reach single-molecule resolution with fluorescence spectroscopy. The interested reader in single-molecule fluorescence techniques and analysis is kindly referred to the papers by Schermelleh et al. [2010], van Mameren et al. [2011] and Yamanaka et al. [2014].

1.3.2 Atomic Force Microscopy

AFM is definitely the most common single-molecule technique because of its traditional application to condensed matter physics since its conception in 1986 [Binning et al., 1986]. AFM uses a cantilever with a very sharp tip to scan surfaces by displacing the sample in the xy plane. During the scanning procedure, the interaction between the surface and the tip bends the cantilever (Figure 1.2a). The reflection of a laser beam focused on the cantilever is captured using a quadrant photodiode (QPD), which is used to measure the deflection with sub-nanometer resolution. In that way, and by assuming cantilever deflections as Hookean, it is possible to calculate the force applied to the sample ($F = -kz$, where k is the stiffness of the

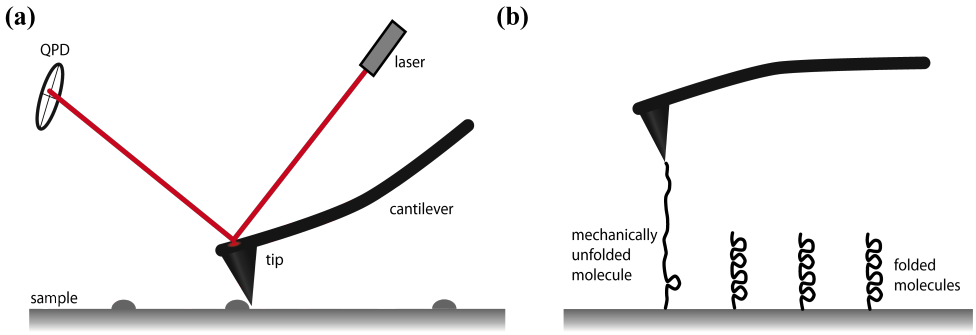

Figure 1.2: Atomic Force Microscopy (AFM). a) A cantilever beam is deflected by the interaction of the tip with the surface. The reflection of a laser on the cantilever surface is measured with a QPD and used as probe of the interactions with the surface. b) AFM in force spectroscopy mode. Retraction of the cantilver mechanically unfolds tip-attached molecule, dropping the force and increasing the extension of the tether.

cantilever and z the deflection). The stiffness of the cantilever determines the range of forces that can be detected and applied, typically spaning from 10 pN to 100 nN. In the standard contact mode, the deflection of the cantilever can be modified by approximating or retracting the sample using a piezoelectric stage. Then, by keeping the deflection/force constant by feedback-loop correction of the distance to the sample, it is possible to obtain a topographic image, which is simply the distance moved by the piezo. However, the AFM can operate with diverse modes of measurement, and the selection of a particular one depends on the final application. For instance, for imaging soft matter samples dynamic modes are preferred. In these modes the tip does not enter into contact with the surface to prevent the damage of the sample, but the cantilever oscillates near its resonance frequency in an atractive regime over the surface. The interaction forces with the sample affect the amplitude and the frequency of the oscillations. Maintaining the amplitude or the frequency constants with a feedback to correct the distance to the surface, a topographic image can be obtained from the displacements of the piezo, similarly as was explained previously for the contat mode [Moreno-Herrero et al., 2004].

AFM can also operate in force-spectroscopy mode. This is an interesting and common aplication of the AFM in biophysics, usually to study biomolecular folding. In AFM force-spectroscopy the molecule of interest is deposited on a surface. Next, a functionalized cantilever tip is brought into contact with the surface to trigger the interaction with the adsorbed molecule. Following retraction of the tip bends the cantilever due to the stiffness of the tethered molecule between the surface and the tip (Figure 1.2b). The folding and unfolding events can be clearly identified as abrupt jumps in the force-distance curve[1] [Rief et al., 1997] where the unfolding distance and the rupture force are a footprint of the structure of the molecule.

[1]In this context, distance indicates the displacement of the piezo.

1.3.3 Optical Tweezers

The first description of optical trapping was published by Arthur Ashkin in 1970 [Ashkin, 1970]. The basic physical principle of optical traps is the radiation pressure, the force appearing by the interaction between electromagnetic fields and matter. In optical tweezers (OT), the reflection and refraction of light from a Gaussian laser beam on a dielectric particle results in a difference of momentum between the incident light and the transmitted light. This difference of linear momentum is transferred to the trapped particle according to Newton's third law, and from the second law this change of momentum per unit of time is the force exerted on the bead. This net force can be decomposed into two forces, the *gradient force* and the *scattering force*. The gradient force points to the centre of the beam and arises because more photons impact on the particle when located on this region due to the Gaussian intensity profile of the beam. Conversely, the scattering force points in the propagation direction of the beam (Figure 1.3a). Therefore, the scattering force needs to be balanced to generate an stable optical trap by preventing the acceleration of the bead in the propagating direction of the beam. This can be achieved by using two counterpropagating lasers or in the most common aproach, using a high numerical aperture objective. This generates a maximum intensity in the focus of the beam, where the gradient force is directed counteracting the scattering force and confining the particle in three dimensions [Ashkin et al., 1986].

During a typical OT experiment, a functionalized molecule is bound on one side to the trapped bead and on the other side, to a reference surface, which can be a glass slide or another bead. In the later case, the position of the reference bead can be fixed by a different and stiffer optical trap or by suction using a micropipette (Figure 1.3b). The displacement of the reference surface generates a force and as a result the molecule is stretched. The applied force can be quantified by considering the trap as a harmonic potential well and thus $F = -kx$, where k is the stiffness of the trap and x is the distance of the confined bead to the center of the trap. Then by using a feedback over the force and the displacement of the surface, it is possible to mantain the force constant. OT allows a wide range of working forces, from 1 pN to 200 pN with subpiconewton resolution [Neuman and Nagy, 2008]. This magnificent resolution in force is possible thanks to the use of a laser-based position detection system. The typical approach consists on the scattering of a laser by the trapped bead and the analysis of the deviations from the equilibrium using a QPD. This methodology results in sub-nanometer spatial resolution and sub-millisecond temporal resolution [Neuman and Block, 2004]. The firsts experiments with OT and biological systems were carried out by Ashkin and colleagues [Ashkin, 1987], but the real revolution arrived with the pioneering experiments by Steven Block, who first reported the movement of kinesin motor proteins along a microtubule [Block et al., 1990]. Four years later, the group of Carlos Bustamante characterized the mechanical properties of a single DNA molecule [Bustamante et al., 1994]. These groups set the basis of a technique that has become a standard for single-molecule biophysics labs. Since then, OT have been extensively used in biophysics to study nucleic acids structure

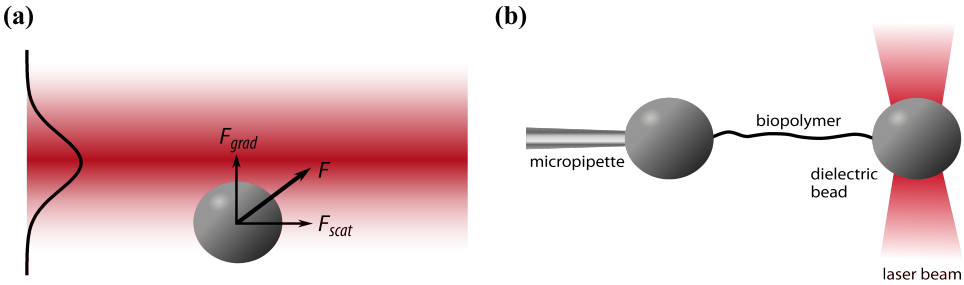

Figure 1.3: Optical Tweezers (OT). a) Ray optics approximation (particle size $\gg$ wavelength) of the optical forces acting on a particle under a collimated Gaussian beam. The total force, F can is decomposed in the gradient force (F_{grad}) pointing towards the gradient of intensity and the scattering force F_{scat} pointing in the propagation direction of the beam. In this configuration the particle is not trapped. b) OT in single beam and pipette configuration. The exerted force can be controlled by translating the position of the pipette while the trap is fixed.

[Bryant et al., 2003], molecular motors [Mallik et al., 2004], DNA processing enzymes [Morin et al., 2015], or protein folding [Stigler et al., 2011].

1.3.4 Magnetic Tweezers

MT were developed several years after the first biological experiments with OT, but its simpler concepcion and its ability to track several molecules in parallel, rapidly positioned them as a practical alternative [Smith et al., 1992; Strick et al., 1996]. MT are based on an inverted microscope focused on a microfluidic chamber where a biopolymer is tethered between the surface of a glass slide and a magnetic bead. In the most common configuration, a couple of magnets located above the sample exert a force on the bead, stretching the tethered molecule (Figure 1.4a). In contrast to OT where the bead is spatially confined, MT are not real *tweezers* as the bead is always pulled vertically by the magnetic force. Using a motorized stage, the magnets can be translated to modify the distance to the bead and in this way it is possible to control the applied force. In addition, magnets can be rotated generating a torque on the bead –and on the molecule– which can be used to characterize the torsional properties of nucleic acids. For a fixed position of the magnets the force is constant, being able to apply forces from 0.01 pN up to 100 pN, although maximum forces are limited by bead size and magnet geometry [Neuman and Nagy, 2008; Lipfert et al., 2009]. Stringent calibrations are necessary for accurate dermination of high forces [Chen et al., 2011]. The applied force can be quantified from the lenght of the tether, L and the variance of the in-plane fluctuations $\langle \delta x^2 \rangle$ as $F = k_B T L / \langle \delta x^2 \rangle$. The coordinates of the molecule are extracted by videomicroscopy with a resolution of a few nanometers (2-10 nm for all axes). Thus, the temporal resolution is limited by the acquisition frequency of the camera, usually to 2-20 ms though faster rates are possible with current CMOS cameras [Otto et al., 2010; Huhle et al., 2015]. An expanded and detailed description of the working principles of MT, including tracking

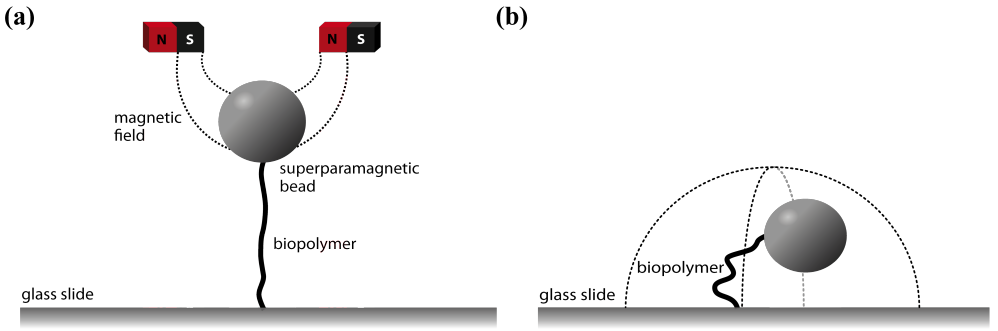

Figure 1.4: Videomicroscopy-based single-molecule techniques. a) Magnetic Tweezers. The molecule is anchored to the surface of a glass slide and to a paramagnetic bead. The application of a non-uniform magnetic field in the z axis generates a force on the bead, extending the tethered molecule. b) Tether Particle Motion. Thermal fluctuations of the bead result in an amplitude of the excursion proportional to the contour length of the molecule.

routines and analytical procedures can be found in the corresponding chapter of this thesis. The possibility of a straighforward application of rotations is one of the major advantages of MT with respect to OT. This ability of exerting torques on the DNA has been critical not only to characterize the response of nucleic acids to supercoiling [Meng et al., 2015] but also to understand proteins involved in the remodeling of the topological state of DNA [Neuman et al., 2009] or the structural effects induced by protein binding [Zorman et al., 2012]. In the following chapters of this thesis, different examples are presented of magnetic tweezers applied to characterize the activity of proteins on the DNA.

1.3.5 Tethered Particle Motion

Despite not being a force spectroscopy technique itself, the TPM technique has been included here due to its simplicity and its similitude in tracking procedures with magnetic and optical tweezers (Figure 1.4b). In TPM experiments, a short DNA molecule ($L < 2000$ bp) is attached to a bead ($R < 1$ µm) in the absence of any applied force and the Brownian fluctuations are measured. The amplitude of the excursions are restricted by the entropic force of the molecule and this is dependent on the lenght and the stiffness of the tethered DNA. Hence, it is possible to track structural changes occuring on the molecule by simply analyzing the fluctuations [Nelson et al., 2006]. In particular, TPM experiments have been widely used to study protein-induced DNA looping dynamics [Chen et al., 2014].

1.3.6 Trends in single-molecule biophysics

Single-molecule techniques are under constant development. During the last years, there have been significant improvements in the field. First, multiplexing has in-

creased the speed of data acquisition, being possible to work with several optical traps simultaneously or with hundreds of beads in parallel in magnetic tweezers [Heller et al., 2017; Cnossen et al., 2007]. Second, the resolution of these techniques has increased dramatically and nowadays it is possible to reach spatial resolutions in the angström range [William et al., 2007] and temporal resolutions of a few microseconds [Neupane et al., 2016; Chung et al., 2015]. And last, a major trend has been the combination of different force-spectroscopy techniques with fluorescence microscopy leading to groundbreaking results on the structural response of the DNA under different conditions [Kemmerich et al., 2016; King et al., 2013; van Loenhout et al., 2012].

Nonetheless, in the last few years there has been not only combinations and improvements of the techniques aforementioned, but additionally, new single-molecule technologies have been developed. This is the case of biological or solid state nanopores [Derrington et al., 2015; Saleh and Sohn, 2003], acoustic force spectroscopy [Sitters et al., 2015] or centrifugal force microscopy [Yang et al., 2016]. It is expected that this trend in the development of new single-molecule techniques and an increase in the number of commercially-available setups will continue in the following years.

1.4 Protein-DNA interactions

The DNA molecule encodes the biological information of the cell in a *four-letters* code. This information stored on the DNA needs to be processed and must be properly organized and copied. However, DNA is not self-sufficient and the cell requires a huge number of proteins to structure, process and replicate this information. Basically, all the functions found in living organisms are dependent on proteins or enzymes, and the synthesis of these is dependent on proteins that are able to interact with the DNA and manipulate it by a wide variety of mechanisms. It then becomes obvious that understanding these processes is crucial to recapitulate the impressive complexity of the cell. The advances of single-molecule techniques are allowing the determination of the working principles of many DNA-binding proteins that are hidden to bulk experiments. For example, single-molecule experiments determined the mechanochemical coupling between ATP hydrolysis and force generation during encapsidation of viral DNA [Chemla et al., 2005] and unveiled the rates of DNA supercoil release and religation by topoisomerases [Strick et al., 2000; Koster et al., 2005]. The expected improvements in single-molecule techniques and analytical methodologies will result in prominent breakthroughs in the comprehension of biological problems involving interactions between DNA and proteins.

1.5 Overview of this thesis

This thesis describes the interaction of two different bacterial proteins with single DNA molecules in the context of DNA replication, segregation and condensation. We have employed magnetic tweezers to understand the implications of forces and torques on these fundamental biological processes. The contents of this thesis have

been organized in three well-separated parts. Part I is a general introduction to the main concepts that will appear recurrently during the rest of the chapters, with Parts II and Part III representing the majority of the experimental results. Lastly, Part IV finishes the dissertation with the main conclusions.

Part I, is an introductory section of the application of magnetic tweezers to DNA molecules and its biological relevance. The **Chapter 2** describes the physical, chemical and biological characteristics of nucleic acids. **Chapter 3**, provides a complete frame of both the instrumental and computational features of a magnetic tweezers apparatus. Both chapters are a useful reference guide for the reader about the concepts exposed in the results sections.

Part II addresses Spo0J-induced condensation of single DNA molecules. First, the system is presented in **chapter 4**. The Spo0J-dependent condensation of the DNA was a novel discovery with important consecuences for understanding the biology of chromosome organization and segregation in Bacteria. In this chapter we present its characterization as well as its dependence with the protein specific sequence *parS*. We find that the specific sequence is not necessary for DNA condensation, and we conclude that condensation is not a consequence of the formation of well-ordered structures. Our results indicate a looping-dependent condensation process. With this idea in mind, we investigated the mechanisms leading to DNA condensation and the protein domains involved on the process. For this purpose we used a manifold of mutants and truncations at different domains of the protein. These results are presented in **chapter 5**. Our results suggest the existence of two well-defined locus for specific and non-specific binding located at the N-terminal and C-terminal domains, respectively.

Part III and its corresponding **chapter 6** deals with the first steps in the rolling-circle replication. In particular, we explored the dynamics of the plasmid replication initiation protein RepC. We found that in order to initiate the rolling-circle replication it is necessary to negatively supercoil the DNA. Our results are in agreement with the requirement of forming a cruciform structure for the activity of RepC. Importantly, in contrast to previously proposed models, our experiments suggest a neglegible role of RepC in the extrusion of the cruciform structure, being this mostly dependent on the supercoiling state of the plasmid.

Finally, **Part IV** summarizes the main conclusions deduced from the work presented in this thesis. Future projects and derived research lines are also included and discussed.

References

Ashkin, A. Acceleration and trapping of particles by radiation pressure. *Phys. Rev. Lett*, 24(4):156–159, 1970.

Ashkin, A. Optical trapping and manipulation of viruses and bacteria. *Science*, 235(4795):1517–1520, 1987.

Ashkin, A., Dziedzic, J. M., Bjorkholm, J. E., and Chu, S. Observation of a single-beam gradient force optical trap for dielectric particles. *Opt. Lett.*, 11(5): 288–290, 1986.

Axelrod, D. Cell-substrate contacts illiminated by total internal reflection fluorescence. *J. Cell Biol.*, 89(1):141–145, 1981.

Binning, G., Quate, C. F., and Geber, C. Atomic force microscope. *Phys. Rev. Lett.*, 56(9):930–933, 1986.

Block, S. M., Goldstein, L. S. B., and Schnapp, B. J. Bead movement by single kinesin molecules studied with optical tweezers. *Nature*, 348(6299):348–352, 1990.

Bryant, Z., D., S. M., Gore, J., Smith, S., R., C. N., and C., B. Structural transitions and elasticity from torque measurements on DNA. *Nature*, 424(6946): 1599–1600, 2003.

Bustamante, C., Marko, J. F., Siggia, E. D., and Smith, S. Entropic elasticity of lambda-phage DNA. *Science*, 265(5175): 1599–1600, 1994.

Bustamante, C., Macosko, J. C., and Wuite, G. J. Grabbing the cat by the tail: manipulating molecules one by one. *Nat. Rev. Mol. Cell. Biol.*, 1(2):721–744, 2000.

Chemla, Y. R., Aathavan, K., Michaelis, J., Grimes, S., Jardine, P. J., L., A. D., and Bustamante, C. Mechanism of force generation of a viral DNA packaging motor. *Cell*, 122(9):683Ű692, 2005.

Chen, H., Fu, H., Zhu, X., Cong, P., Nakamura, F., and Yan, J. Improved high-force magnetic tweezers for stretching and refolding of proteins and short DNA. *Biophys. J.*, 100(2):517–523, 2011.

Chen, Y.-J., Johnson, S., Mulligan, P., Spakowitz, A. J., and Phillips, R. Modulation of DNA loop lifetimes by the free energy of loop formation. *Proc. Natl. Acad. Sci. U.S.A.*, 111(49):17396–17401, 2014.

Chuang, H.-Y., Hofree, M., and Ideker, T. A decade of systems biology. *Annu. Rev. Cell Dev. Biol.*, 26:721–744, 2010.

Chung, H. S., Piana-Agostinetti, S., Shaw, D. E., and Eaton, W. A. Structural origin of slow diffusion in protein folding. *Science*, 349(6255):1504–1510, 2015.

Cnossen, J. P., Dulin, D., and Dekker, N. H. An optimized software framework for real-time, high-throughput tracking of spherical beads. *Rev. Sci. Instrum*, 85:257–278, 2007.

Derrington, I. M., Craig, J. M., Stava, E., Laszlo, A. H., Ross, B. C., Brinkerhoff, H., Nova, I. C., Doering, K., Tickman, B. I., Ronaghi, M., Mandell, J. G., Gunderson, K. L., and Gundlach, J. H. Subangstrom single-molecule measurements of motor proteins using a nanopore. *Nat. Biotechnol.*, 33(10): 1073–1075, 2015.

Heller, I., Laurens, N., Vorselen, D., Broekmans, O. D., Biebricher, A. S., King, G. A., Brouwer, I. Wuite, G. J. L., and Peterman, E. J. Versatile quadruple-trap optical tweezers for dual DNA experiments. *Methods Mol. Biol.*, 1486: 257–278, 2017.

Huhle, A., Klaue, D., Brutzer, H., Daldrop, P., Joo, S., Otto, O., Keyser, U. F., and Seidel, R. Camera-based three-dimensional real-time particle tracking at khz rates and angstrom accuracy. *Nat. Commun.*, 6:5885, 2015.

Karplus, M. and McCammon, A. J. Molecular dynamics simulations of biomolecules. *Nat. Struct Biol.*, 9(10): 646–652, 2002.

Kelley, A. M., Michalet, X., and Weiss, S. Single-molecule spectroscopy comes of age. *Science*, 292(5522):1671–1672, 2001.

Kemmerich, F. E., Swoboda, M., Kauert, D. J., Grieb, M. S., Hahn, S., Schwarz, F. W., Seidel, R., and Schlierf, M. Simultaneous single-molecule force and fluorescence sampling of DNA nanostructure conformations using magnetic tweezers. *Nano Lett.*, 16(1):381–386, 2016.

King, G. A. amd Gross, P., Bockelmann, U., Modesti, M., Wuite, G. J., and Peterman, E. J. Revealing the competition between peeled ssDNA, melting bubbles, and S-DNA during DNA overstretching using fluorescence microscopy. *Proc. Natl. Acad. Sci. U.S.A.*, 110(10):3859–3869, 2013.

Koster, D. A., Croquette, V., Dekker, C., and Shuman, S. Dekker, N. H. Friction and torque govern the relaxation of DNA supercoils by eukaryotic topoisomerase IB. *Nature*, 434(7033):671–674, 2005.

Lipfert, J., Haom, X., and Dekker, N. H. Quantitative modeling and optimization of magnetic tweezers. *Biophys. J.*, 96(12):5040–5049, 2009.

Lotka, A. J. Analytical note on certain rhythmic relations in organic systems. *Proc. Natl. Acad. Sci. U.S.A*, 6(2):31–113, 1920.

Mallik, R., Carter, B. C., Lex S. A. King, S. J., and Gross, S. P. Cytoplasmic dynein functions as a gear in response to load. *Nature*, 427(6975):649–652, 2004.

McCammon, J. A., Gelin, B. R., and Karplus, M. Dynamics of folded proteins. *Nature*, 267(5612):585–90, 1977.

Meng, H., Bosman, J., van der Heijden, T., and van Noort, J. Coexistence of twisted, plectonemic, and melted DNA in small topological domains. *Biophys. J.*, 105(5):1174–1181, 2015.

Milo, R. and Philips, R. *Cell Biology by the Numbers*. Garland Science, 1st edition, 2015.

Moreno-Herrero, F., Colchero, J., Gomez-Herrero, J., and Baró, A. M. Atomic force microscopy contact, tapping, and jumping modes for imaging biological samples in liquids. *Phys. Rev. E. Stat. Nonlin. Soft. Matter Phys.*, 69(3): 1–8, 2004.

Morin, J. A., Cao, F. J., Lázaro, J. M., Arias-Gonzalez, J. R., Valpuesta, J. M., Carrascosa, J. L., Salas, M., and Ibarra,

B. Mechano-chemical kinetics of replication: identification of the translocation step of a replicative DNA polymerase. *Nucleic Acids Res.*, 43(7):3643–3652, 2015.

Nelson, P. C., Zurla, C., Brogioli, D., Beausang, J. F., Finzi, L., and Dunlap, D. Tethered particle motion as a diagnostic of DNA tether length. *J. Phys. Chem. B*, 110(34):17260–17267, 2006.

Neuman, K. C. and Block, S. M. Optical trapping. *Rev. Sci. Instrum.*, 75(9):2787–2809, 2004.

Neuman, K. C. and Nagy, A. Single-molecule force spectroscopy: optical tweezers, magnetic tweezers and atomic force microscopy. *Nat. Methods*, 5(6):491–505, 2008.

Neuman, K. C., Charvin, G., Bensimon, D., and Croquette, V. Mechanisms of chiral discrimination by topoisomerase IV. *Proc. Natl. Acad. Sci. USA*, 106(14):6986–91, 2009.

Neupane, K., Foster, D. A., Dee, D., Yu, H., Wang, F., and Woodside, M. T. Direct observation of transition paths during the folding of proteins and nucleic acids. *Science*, 352(6282):239–242, 2016.

Otto, O., Czerwinski, F., Gornall, J. L., Stober, G., Oddershede, L. B., and Seidel, R. Keyser, U. F. Real-time particle tracking at 10,000 fps using optical fiber illumination. *Opt. Express.*, 18(22):22722–33, 2010.

Rief, M., Gautel, M., Oesterhelt, F., Fernández, J. M., and E., G. H. Reversible unfolding of individual titin immunoglobulin domains by AFM. *Science*, 276(5315):1109–12, 1997.

Ritort, F. Single-molecule experiments in biological physics: methods and applications. *J Phys. Condens. Matter.*, 18 (35):531–583, 2006.

Roy, R., Hohng, S., and Ha, T. A practical guide to single-molecule FRET. *Nat. Methods*, 5(6):507–516, 2008.

Saleh, O. A. and Sohn, L. L. An artificial nanopore for molecular sensing. *Nano Letters*, 3(1):37–38, 2003.

Schermelleh, L., Heintzmann, R., and Leonhardt, H. A guide to super-resolution fluorescence microscopy. *J. Cell Biol.*, 190(2):165–175, 2010.

Schrödinger, E. *¿Qué es la vida?* Tusquets Editores, 3rd edition, 1993. ISBN 9788472236073.

Sitters, G., Kamsma, D., Thalhammer, G., Ritsch-Marte, M., Peterman, E. G. J., and Wuite, G. J. L. Acoustic force spectroscopy. *Nat. Methods.*, 12(1):47–50, 2015.

Smith, S. B., Finzi, L., and Bustamante, C. Direct mechanical measurements of the elasticity of single DNA molecules by using magnetic beads. *Science*, 252 (5085):1122–6, 1992.

Stigler, J., Ziegler, F., Gieseke, A., Gebhardt, J. C., and Rief, M. The complex folding network of single calmodulin molecules. *Science*, 334(6055):512–516, 2011.

Strick, T. R., Allemand, J. F., Bensimon, D., Bensimon, A., and Croquette, V. The elasticity of a single supercoiled DNA molecule. *Science*, 271(5257):1835–1837, 1996.

Strick, T. R., Croquette, V., and Bensimon, D. Single-molecule analysis of DNA uncoiling by a type II topoisomerase. *Nature*, 404(6780):901–904, 2000.

Turing, A. M. The chemical basis of morphogenesis. *Phil. Trans. R. Soc. B*, 237(641):37–72, 1952.

Vafabakhsh, R. and Ha, T. Extreme bendability of sub-100 bp long DNA revealed by single molecule cyclization. *Science*, 337(6098):1097–1101, 2013.

van Loenhout, M. T. J., de Grunt, M. V., and Dekker, C. Dynamics of DNA supercoilsg. *Science*, 338(6103):94–97, 2012.

van Mameren, J., Wuite, G. J. L., Heller, I., and G., P. E. J. *Single Molecule Analysis: Methods and Protocols.* Methods in Molecular Biology 783. Humana Press, 1 edition, 2011.

Volterra, V. Variazioni e fluttuazioni del numero d'individui in specie animali conviventi. *Mem. Acad. Lincei Roma*, 2: 31–113, 1926.

William, G. J., Woodside, M. T., and Block, S. M. High-resolution, single-molecule measurements of biomolecular motion. *Annu. Rev. Biophys.*, 36:171–190, 2007.

Yamanaka, M., Smith, N. I., and Fujita, K. Introduction to super-resolution microscopy. *Microscopy*, 63(3):177, 2014.

Yang, D., Ward, A., Halvorsen, K., and Wong, W. P. Multiplexed single-molecule force spectroscopy using a centrifuge. *Nat. Commun.*, 7:47–50, 2016.

Yasuda, R., Noji, H., Yoshida, M., Kinosita, K. J., and Itoh, H. Resolution of distinct rotational substeps by submillisecond kinetic analysis of F1-ATPase. *Nature*, 410(6831):898–904, 2001.

Zorman, S., Seitz, H., Sclavi, B., and Strick, T. R. Topological characterization of the DnaA-oriC complex using single-molecule nanomanipulation. *Nucleic Acids Res.*, 40(15):7375–7383, 2012.

2

Nucleic acids

Chapter overview. Nucleic acids were discovered during the second half of the 19th century. It took 80 years to determine their chemical composition and 100 more to determine their characteristic structure. Today, they are still a collaborative and meeting point for physicists, chemists and biologists. The purpose of this chapter is to describe the basis of nucleic acids' chemical and physical features and their implications for the living organisms.

2.1 Introduction

Nucleic acids were discovered by the Swiss biologist Friedrich Miescher in 1868. The term *nucleic acids* was later selected because they were detected by optical microscopy as a dense matter located in the nucleous of the cell when stained with basic dyes. The affinity for basic dyes revealed its acidic nature, and indicated that they were negatively charged, providing the first inferences about its chemical nature. Curiously, in 1865, only three years before the discovery of nucleic acids, Gregor Mendel had published *Versuche über Pflanzen-Hybriden* [Mendel, 1866], where the inheritance of biological characteristics was formally analyzed, setting the basis of genetics. In 1902 Theordor Boveri and Walter Sutton independently proposed *The chromosome theory*, suggesting that the inheritance factors –at the time known as Mendel's factors and nowadays as genes– were located on the chromosomes. The theory was later probed in 1911 by Thomas Hunt Morgan [Klug et al., 2005; Griffiths et al., 2002]. Nonetheless, it was not clear whether the nucleic acids present on the chromosomes were the information molecules or simple structural components for non-discovered *information proteins*. Avery-MacLeod-McCarty experiments in 1944 and the Hersey-Chase experiment in 1952 definitively confirmed that nucleic acids –and not the proteins– were the molecules bearing the genetic information of the organisms [Watson et al., 2014]. It is not surprising that the relation established between a physical molecular entity and inheritance lead to an increasing interest on the study of nucleic acids. This interest, both from a biological and a chemical point of view, was impeled by the discovery of the double-helix structure of the DNA by James D. Watson and Francis Crick in 1953, and boosted molecular biology as scientific discipline during the second half of the 20th century.

2.2 DNA Structure

2.2.1 Primary structure

Nucleic acids are biopolymers composed of monomers denominated nucleotides. Nucleotides are organic molecules formed by nucleosides bonded through a phosphate backbone. Nucleosides is the term used for a molecule consisting of a nitrogen base and a five-carbon sugar [Sinden, 1994]. In the following points the chemical structure of each component is expanded.

Nitrogen bases

The bases are aromatic compounds derived from purine and pyrimidine. Purines are heterocyclic rings and in the nucleic acids can be found in the form of Adenine (A) or Guanine (G) (Figure 2.1a). Adenine presents an amino group (-NH_2) in C6 while guanine has an amino group at C2 and a carbonyl group (-CO-) at C6. Pyrimidine bases are composed of a six-members ring (Figure 2.1b). Thymine (T) and cytosine (C) are found in DNA and they differ in the positions C4 and C5. Thymine has a carbonyl group in C4 and a methyl group at C5, while cytosine has an amino group at C4 and a hydrogen at C5. Uracil (U) is chemically more similar to thymine, but the methyl group of thymine at C5 is replaced by a hydrogen. The similarity of uracil and thymine is reflected in the role of uracil substituting thymine in ribonucleic acid (RNA), while it is almost absent in deoxyribonucleic acid (DNA).

Nitrogen bases are planar structures that can interact within them forming hydrogen bonds. The most energetically favorable form of interaction is between a purine and a pyrimidine. Particularly, guanine interacts with cytosine through the formation of three hydrogen bonds, and adenine interacts with thymine or uracil by two hydrogen bonds. These bondings denominated Watson-Crick base-pairs stabilize the double-stranded structure described in the following section.

Sugar

There are two sugars as components of nucleic acids, β-D-ribose and β-D-deoxyribose (Figure 2.2). Both are derived from the D isomer of ribose –a pentose monosaccharide– and, in its ring form, are constituents of nucleic acids. β-D-ribose is the sugar present in RNA, while in DNA the pentose is β-D-deoxyribose, whose only difference with β-D-ribose is the absence of the hydroxyl group at the 2' posi-

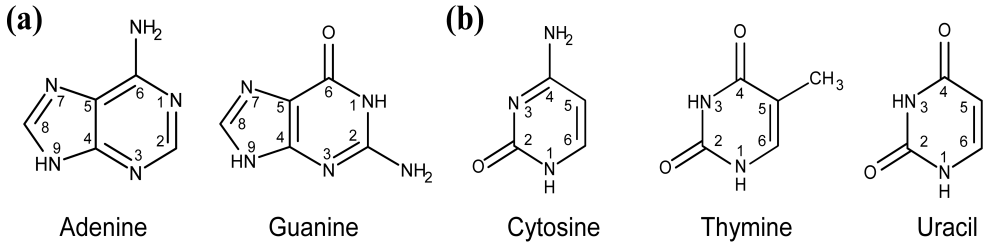

Figure 2.1: Chemical structure of the nitrogen bases. a) Purine bases adenine and guanine. b) Pyrimidine bases thymine, cytosine and uracil. Numbering of the rings is indicated.

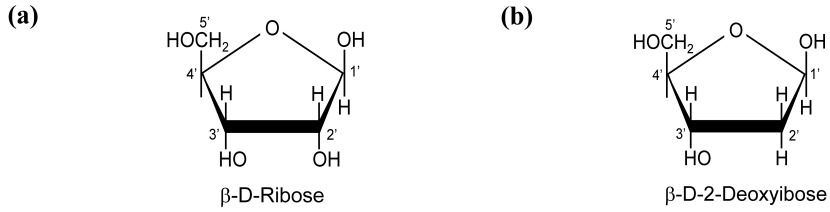

Figure 2.2: Chemical structure of RNA and DNA pentoses in Haworth projection. a) β-D-Ribose is the sugar constituting the RNA. b) β-D-2-Deoxyribose is the constituting of DNA. Its difference with β-D-Ribose relies in the substitution of the hydroxy group by hydrogen at the 2' carbon. The numbering system of the carbon in the sugar is indicated with number and prime.

tion.

The pentose is a non-planar and very flexible ring [Saenger, 1988; Sinden, 1994], and adopts different conformations with important implications for the global structure and flexibility of duplex nucleic acid molecules [Marin-Gonzalez et al., 2017].

Nucleosides and nucleotides

Nucleosides result from the glycosidic linkage between a nitrogen base and the sugar (Figure 2.3a). The bond occurs between the C1' of the pentose and the amino group in the position 9 (purines) or in position 1 (pyrimidines).

Nucleotides are nucleosides with phosphate groups in the C5' position of the pentose. In the living organisms, nucleotides are commonly synthesized in a triphosphate form, where each phosphate is named from α to γ starting from the C5' phosphate (Figure 2.3b). A paradigmatic example of a triphosphate nucleotide is the ATP molecule, the *energy coin* of biological reactions.

The synthesis of DNA and RNA is dependent on triphospate nucleotides. The polymerization reaction occurs through the formation of a phosphodiester bond between the 3'OH in the terminal nucleotide of the polymerized strand and the 5' phosphate in α of the entering triphosphate nucleotide, releasing a pyrophosphate group $P_2O_7^{4-}$ and a H_2O molecule (Figure 2.3c). Thus, DNA synthesis proceeds from 3'$\rightarrow$5': the resulting strand presents a 3'OH group at one end and a 5'PO_4^{3-} at the other end and, as such, nucleic acids are molecules with a well-defined direction.

2.2.2 Secondary and tertiary structures

The DNA forms a double helix structure discovered by Watson and Crick in 1953. The double helix is formed between two antiparallel strands. In other words, one strand oriented in 5'$\rightarrow$3' pairs with a strand in 3'$\rightarrow$5', and thus at each end of the helix there is a 5' end and a 3' end at opposing strands. The double helix structure is maintained by the Watson-Crick base pairing described previously and by stacking interactions. These correspond to the hydrophobic interaction between the rings of adjacent bases of the same strand as well as Van der Waals interactions of the induced dipole on the ring [Sinden, 1994].

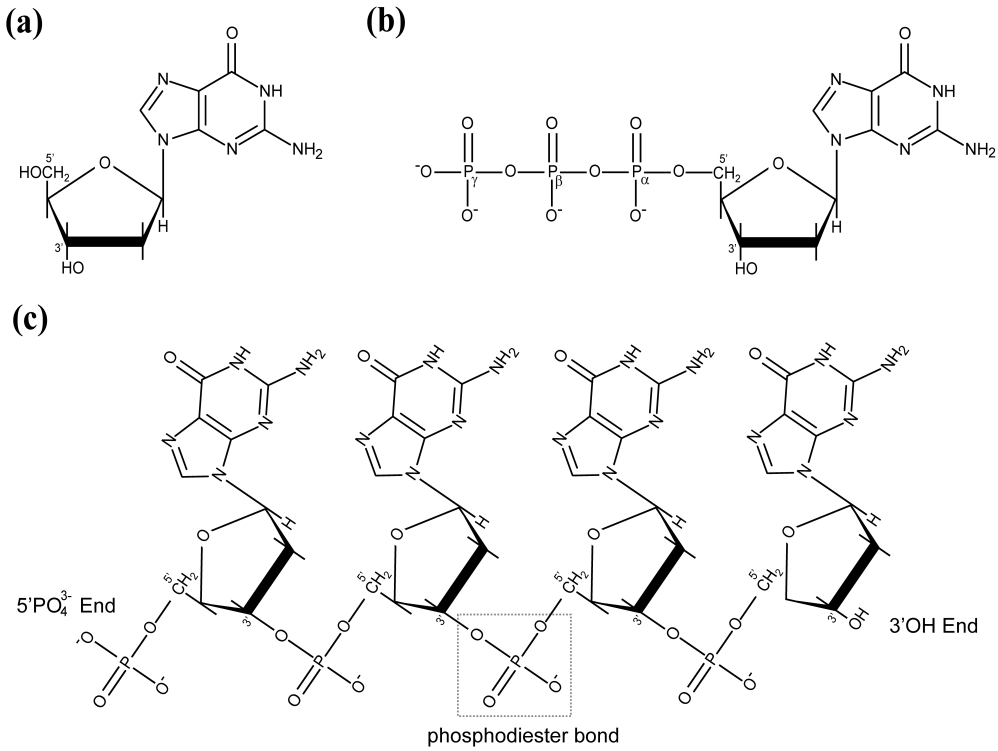

Figure 2.3: Chemical structure of nucleosides, nucleotides and a DNA strand. a) The nucleoside Deoxyguanosine. b) The nucleotide deoxyguanosine 5'-triphosphate (dGTP). c) Structure of a poly-G DNA strand highlighting the phosphodiester bond. The polarization of the molecule due the 5' PO_4^{2-} and 3'OH ends is indicated.

The DNA arranges in a double-helix structure termed B-DNA. This is the most common structure for DNA under physiological conditions (Figure 2.4a). The B-DNA structure is a right-handed double helix with a diameter of 2 nm. The helix has an inter base-pair distance of 0.34 nm, and 3.4 nm per turn, or approximately 10.5 bp per turn. Remarkably, the double helix is not perfectly symmetric, but presents two different grooves of different size, a major groove and a minor groove (Figure 2.4b). This distinct structural feature results in specificity for the binding of proteins and drugs to one or other groove, as will be discussed later.

The B form is not the only helical structure found in nucleic acids. There are alternative helixes that can be formed under specific conditions of salt and humidity [Saenger, 1988]. Among them, the better studied is the so-called A-form[1]. The A structure is a right-handed helix with a diameter of 2.3 nm and a helical pitch of 2.8 nm, wider and shorter than the B form (Figure 2.4c). This is a consequence, among others, of a reduced distance between adjacent base-pairs to 0.255 nm.

A-DNA was originally described from crystals in the presence of ethanol, which

[1]This terminology was given by Rosalind Franklin, crystalographer and co-discoverer of the B-DNA structure.

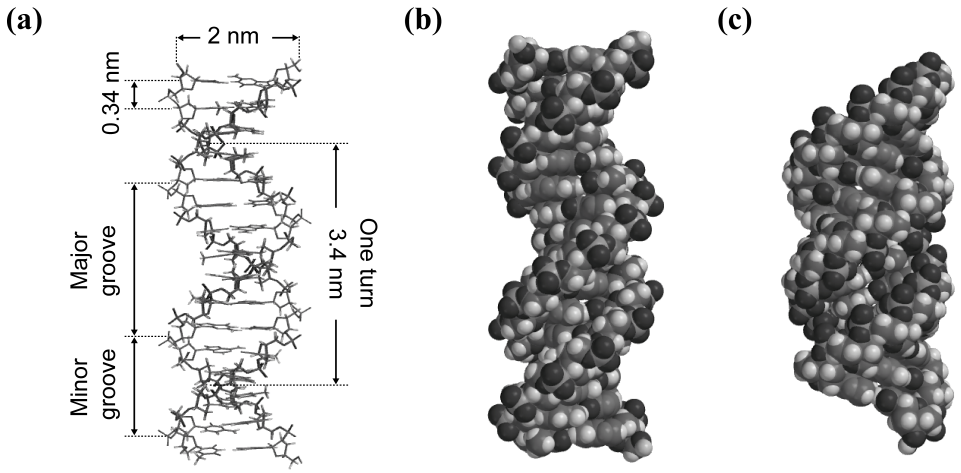

Figure 2.4: Double helix structure of nucleic acids. a) Structural parameters of B-DNA. b) Space filling structure of 12-mer B-DNA clearly shows a thinner diameter and a longer extension than the same molecule in its A-form shown in pannel (c). Structures generated with Spartan (Wavefunction, Inc).

was suggested to induce B → A. transition. However, the formation of A-DNA in ethanolic solutions is still under debate [Hormeno et al., 2011]. It has been proposed that the ethanol-induced B → A transition is only possible in regions enriched in repeats of G or C poly(dG-dC). Remarkably, RNA, which can be found *in vivo* in a double-stranded form, as well as hybrids of DNA-RNA which are also present in living organisms during transcription, adopt an A structure [Saenger, 1988]. Therefore, A-form helices are not just *in vitro* curiosities but they have important biological implications.

2.3 Quaternary structure: DNA Supercoiling

The DNA inside the cell of all living organisms is constrained on its ends. This can be due to its circular nature or due to the attachment to proteinaceous and/or cellular scaffolds in linear DNA molecules [Calladine and Drew, 1997]. In this scenario of constrained ends, also known as torsionally-constrained molecules, exerting torques on the DNA results in the formation of *supercoil* structures similar to that adopted by the cord of old telephones (Figure 2.5a). These structures are known as plectonemes but other supercoiling structures can be formed with a solenoidal arrangment (Figure 2.5b). The genome is subjected to forces and torques exerted by the DNA-processing machines, and hence supercoils naturally exist in the cell in the form of plectonemes (Figure 2.5c) or in the form of solenoidal turns. In fact, suppercoiling plays an important role for the cell, being involved in the organization and expression of the genetic material and therefore being strictly regulated.

The supercoiling of the DNA can be described by a mathematical formalism derived by White [1969] and applied to circularly-closed DNA by Fuller [1971]. In White's description, an important topological number is the *linking number*, Lk

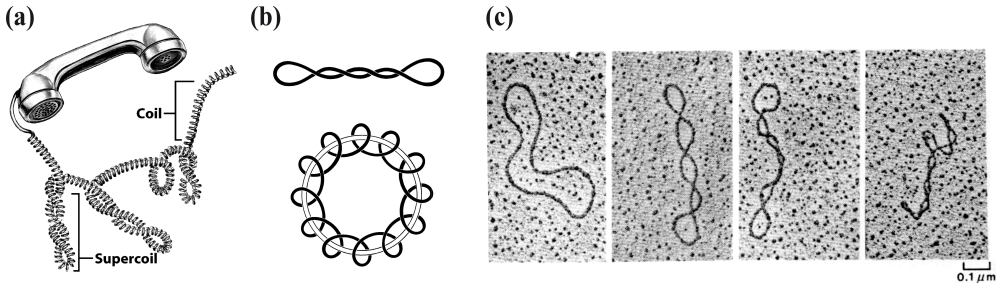

Figure 2.5: Supercoiling is a naturally-occurring phonomenon. a) Supercoils in telephone chords. Adapted from Cox and Nelson [2008]. b) Comparison between plectonemic supercoils (*up*) and solenoidal supercoils (*down*). Adapted from Calladine and Drew [1997]. c) Electron micrographs of a 5 kb plasmid with supercoiling degrees increasing from left (relaxed) to right showing plectonemic structures. Adapted from Kornberg and Baker [1992].

which is an integer value defined as the number of times one strand crosses the other given by

$$Lk = Tw + Wr \tag{2.1}$$

In this expression, Tw is the *twist*, the number of crossings of one strand with respect to the other. The *writhe*, Wr states for the number of times the central axis of the double helix crosses with itself, and as such represents the number of plectonemic or solenoidal turns on the DNA.

Relaxed dsDNA has a $Lk > 0$ given by the natural twist of the double helix i.e $Lk_0 = Tw_0 = L/h$, where L is the length of the DNA and h is the helical pitch (Figure 2.6a). For instance, a 210 bp dsDNA molecule has $Tw_0 = 20$ and consequently it is necessary to apply 20 negative turns to obtain two paired linear dsDNA molecules ($Lk = 0$). Linear molecules with free ends or molecules, either linear or circular, with broken phosphodiester bonds –known as nicks in the nucleic acids field– also present this value of Tw_0, but the linking number is not defined in these situations. Nicked DNA molecules always have $Wr = 0$, since the non-nicked strand can rotate at the nicking position relaxing the superimposed torsional stress.

The value of Lk is a topological invariant i.e. it is a constant as long as the backbone of the strands remains continuous. In the case of dsDNA, this implies the absence of nicks for both strands. Contrary, Tw and Wr are not invariants and can be modified acquiring non-integer values. Nonetheless, since $Lk = \text{const.}$, changes in twist are compensated by changes of writhe (or twist if writhe is the altered parameter). In mathematical terms, $\Delta Tw = -\Delta Wr$ (Figure 2.6b and Figure 2.6c). This relation holds as long as writhe supercoils are allowed to form. If the formation of writhe is constrained by an external source, for instance an stretching force, the linking number of the molecule is no longer constant $Lk \neq Lk_0$ (Figure 2.6c). The same happens if twist could not be modified.

The DNA is considered in a supercoiled state when $\Delta Lk = Lk - Lk_0 \neq 0$. Then,

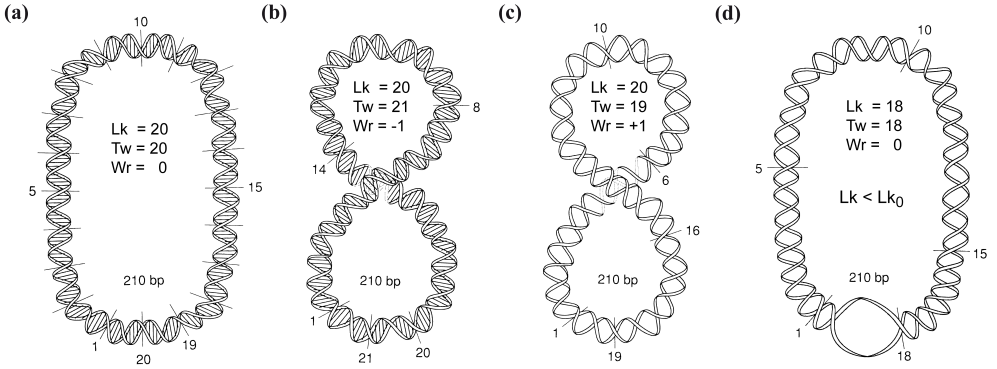

Figure 2.6: Relationship of the linking number (Lk) with twist (Tw) and writhe (Wr) in covalently-closed DNA molecules. a) Relaxed DNA molecule with 210 bp, corresponding to $Lk_0 = Tw_0 = 20$ turns. b) The introduction of a negative writhe turn $Wr = -1$ is counterbalanced with a positive twist turn, keeping Lk constant. c) Introduction of a negative twist turn results in the formation of a postive writhe turn $Wr = +1$. d) The formation of a 21 bp melting bubble results in the unwinding of 2 turns ($\Delta Tw = -2$). If the formation of plectonemes is prevented, ΔWr =0 and therefore the DNA is in negative supercoiling state $Lk < Lk_0$. Adapted from Sinden [1994].

is practical to define the relative supercoiling with respect to the relaxed form,

$$\sigma = \frac{\Delta Lk}{Lk_0}. \tag{2.2}$$

The term *supercoiling degree* is used for the parameter σ. For $\sigma > 0$ the DNA is positively supercoiled. If DNA is positively supercoiled and $Wr = 0$ the DNA is overwound. That is to say, it has less base pairs per turn of the double helix. The reader is invited to deduce this result. For $\sigma < 0$ the DNA is negatively supercoiled. In this case, for $Wr = 0$ the DNA is underwound and hence it has more base pairs per turn. In the cell, the DNA is usually negatively supercoiled, with the exception of thermophilic organisms. The supercoiling of the DNA is a physical feature that is exploited to package the genome inside the cell.

In prokaryotes, DNA is negatively supercoiled by writhe turns adopting plectonemic structures (Figure 2.6c). The negative supercoiling of the DNA is achieved by the activity of DNA gyrase, an ATP-dependent topoisomerase[2]. DNA gyrase uses the energy of ATP to produce a nick in both strands to introduce two negative writhe turns, $\Delta Wr = -2$, without altering twist, $\Delta Tw = 0$, and accordingly $\Delta Lk = \Delta Wr = -2$. The action of DNA processing machines, as well as topoisomerases relieving supercoiling, results in an equilibrium in the supercoiling degree of the cell, and this has been determined to be $\sigma \approx$ -0.05 [Wang, 1984]

In eukaryotic cells, the DNA is not supercoiled forming plectonemes, but solenoid-like structures after binding to histones. Histones are small proteins involved in the organization of the eukaryotic chromosome by wrapping two DNA turns around its structure. Since histone binding is linked to a reduction of Tw and due to the

[2]Proteins involved in the regulation of DNA supercoiling.

activity of topoisomerases, the overall change in linking number per histone bound is $\Delta Lk = -1$. The change of linking number is concerted to the formation of a negative writhe turn $\Delta Wr = -1$ in the wrapped DNA.

2.4 Polymer models

Polymers in a thermal bath acquire a configuration that maximizes their entropy by means of bending fluctuations. As a result, the distance between the ends of the chain does not equal the length of the polymer, depending on the stiffness of the molecule. Force-spectroscopy techniques essentially measure molecular extension as a result of an applied and controlled tension. The combination of experimental results with theoretical models is desired to describe and predict the response in terms of extension as a function of the force applied. In this section, two widely used models are briefly described.

2.4.1 Kratky-Porod and Freely-jointed chain models

The Kratky-Porod model (KP) considers a polymer as a set of N ideal (non self-avoiding) stiff segments of equal length b and direction vector $\hat{\mathbf{t}}$ [Strick et al., 2000]. The bending energy U_b of this chain is given by

$$\begin{aligned} U_b &= -\frac{B}{b}\sum_{i=2}^{N}\hat{\mathbf{t}}_i\cdot\hat{\mathbf{t}}_{i-1} \\ &= -\frac{B}{b}\sum_{i=2}^{N}\cos\theta_i, \end{aligned} \tag{2.3}$$

where B is the bending modulus and θ_i is the angle between direction vectors $\hat{\mathbf{t}}_i$. Under the action of an stretching force, the work done on the segment U_s is given by

$$\begin{aligned} U_s &= -Fb\sum_{i=1}^{N}\hat{\mathbf{t}}_i\cdot\hat{\mathbf{u}}_{\mathbf{F}} \\ &= -Fb\sum_{i=1}^{N}\cos\phi_i. \end{aligned} \tag{2.4}$$

In the former expressions, $\hat{\mathbf{u}}_{\mathbf{F}}$ states for the unitary vector in the direction of the applied force and ϕ_i is the angle between the ith segment and the stretching force axis.

Combining the equation for the bending and stretching energies, results in a total energy $U = U_b + U_s$. An important approximation results when considering $U_b = 0$. This model is known as Freely-jointed chain (FJC) and can be solved analytically to determine the expected extension l as a function of the stretching force. The average extension at force F can be determined as

$$l = L\left(\coth\frac{Fb}{k_BT} - \frac{k_BT}{Fb}\right) \tag{2.5}$$

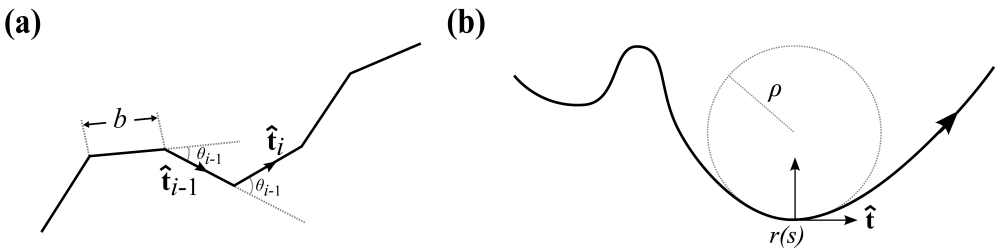

Figure 2.7: Polymer models. a) Discrete model. The polymer is modellized as N non-interacting segments of lenght b with orientation vector $\hat{\mathbf{t}}$. b) Continuous approximation. The worm-like chain model (WLC) considers the configuration of the polymer as a curve $r(s)$. At a certain arc length s, the curvature κ is the inverse of the radius of curvature ρ and dependent on the tangent to the curve $\hat{\mathbf{t}}(s)$.

Here b is also known as *Kuhn length* related with the bending modulus as $b = 2B/k_BT$, and L is the total length of the chain, known in the field as *contour lenght*. A complete derivation of this expresion can be found in the Appendix A.

This approximate model does not capture accurately the response of dsDNA at the range of forces typically used in single-molecule experiments. However, it works well for ssDNA and has been widely used to characterize its mechanical response [Dessinges et al., 2002; Bosco et al., 2014].

2.4.2 Worm-like chain model

The worm-like chain (WLC), can be understood as the continuous version of the KP model by considering $b \to 0$. The configuration acquired by polymer can be described as a smooth curve $\vec{r}$ parameterized as a function of the arc lenght s, $\vec{r} = \vec{r}(s)$. It is possible to define a tangent unitary vector along the curve $\hat{\mathbf{t}}(s)$ as

$$\hat{\mathbf{t}}(s) = \frac{d\vec{r}}{ds}. \tag{2.6}$$

At a certain position $\vec{r}(s)$ the curvature κ is given by

$$\kappa(s) = \left| \frac{d\hat{\mathbf{t}}(s)}{ds} \right|, \tag{2.7}$$

and this equals the inverse of the radius of curvature ($\rho(s) = 1/\kappa(s)$) and the angular change $d\theta$ between adjacent infinitesimal segments [Adams, 2009],

$$\kappa(s) = \frac{d\theta}{ds}. \tag{2.8}$$

In order to calculate the bending energy of such polymer, the bending energy for a couple of segments is considered to follow a harmonic dependency with the angle, $U = \frac{1}{2}B\theta^2$ [Bloomfield et al., 1999]. Therefore, the total energy of the chain is

$$U_b = \frac{1}{2}B \int_0^L \kappa^2(s) ds, \tag{2.9}$$

where L is the contour length of the chain [Nelson, 2007; Phillips et al., 2009]. The *spring constant* equals the bending modulus and this is given by $B \equiv k_B T\xi$, where ξ is the *persistence length*, a measurement of the angular correlation between adjacent segments.

In the presence of tension, bending fluctuations are constrained by the exerted force. The total energy U of the polymer includes a stretching term U_s and is given by $U = U_b - U_s$. This stretching term under the action of a force F reads

$$U_s = F \int_0^L \hat{\mathbf{t}}(s) \cdot \hat{\mathbf{u}}_\mathbf{F}\, ds \tag{2.10}$$

where $\hat{\mathbf{u}}_\mathbf{F}$ is a unit vector in the direction of the applied force F. Marko and Siggia derived an interpolation formula to describe the expected force for a molecule at a certain extension [Marko and Siggia, 1995],

$$F = \frac{k_B T}{\xi}\left[\frac{1}{4(1-l/L)^2} - \frac{1}{4} + \frac{l}{L}\right]. \tag{2.11}$$

Alternatively, it is algebraically more convenient to use the expression proposed by Moroz and Nelson [Moroz and Nelson, 1997] from the approximations by Marko and Siggia [1995] and Odijk [1995] to determine the extension at a given stretching force,

$$\frac{l}{L} = 1 - \frac{1}{2}\left[\frac{F\xi}{k_B T} - \frac{1}{32}\right]^{-1/2}, \tag{2.12}$$

which is valid for $F > 0.1$ pN. Both equations have been proved to accurately describe experimental data and are routinely used to fit force-extension curves from single-molecule experiments.

2.5 Nucleic acid-protein interactions

The interaction between nucleic acids and proteins is intrinsic to the existence of life, at least considering *life* as is known today on Earth. According to the RNA world hypothesis, life was originally protein-independent and nucleic acids were both information carriers and catalytic elements [Lazcano and Miller, 1996]. During the evolutionary process, Nature selected proteins as catalytic macromolecules due their higher chemical flexibility, while nucleic acids were selected as the source of genetic information. This evolutionary history resulted in a dependence of the interaction of proteins with nucleic acids to transduce and regulate the information storaged on the laters. In addition, at the present time, there are still strong interactions between nucleic acids and proteins forming a stable nucleoprotein complex with specific catalytic activities [Cech, 2009]. The fundamental example of nucleoprotein is the ribosome, but also the CRISPR/Cas system and telomerase enzymes, which are very active research topics. This section as well as this thesis are, however, focused on the interaction between DNA and proteins as independent elements.

As it has been mentioned previously, the information storage on the DNA requires to be copied, read, mantained and structured to ensure the survival of the organism

and its offspring. These tasks are carried by a vast number of DNA-processing proteins coordinating their activities in space and time. Surprisingly, despite the big variability in functions and sequences recognized by DNA binding proteins, there are certain *building blocks* common in the structure of DNA-binding proteins. In this section a brief introduction to the most common structural motifs involved in DNA recognition and its structural implications are presented.

2.5.1 Non-specific interactions

Non-specific binding is mostly dependent on the interaction with the pentose and the phosphate backbone. The phosphates provide a homogeneous negatively-charged surface that can be recognized by positively charged aminoacids of the protein. A textbook example of non-specific DNA binding proteins are the already mentioned histones. Histones are abundant proteins found in all eukaryotes and are involved in the condensation of the genome. These proteins wrap non-specific DNA by the electrostatic interactions between the positively charged residues lysines and arginine of the histone and the negatively charged phosphate backbone.

2.5.2 Specific interactions

The specific interaction of proteins with the DNA implies the recognition of particular DNA sequences. In some cases, the specific interaction with a given sequence lies on the particular structure adopted by the DNA due the structural peculiarities of the sequence rather than in the recognition of the base pairs [Rohs et al., 2010]. This is the case of cruciform structures or highly-bent regions. However, the most common recognition mechanism of specific DNA sequences relies on the recognition of the base pairs, although this is usually coordinated with subtle sequence-dependent structural effects [Sinden, 1994]. Thus, it is not surprising that most specific protein-DNA interactions occur on the major groove, where the base pairs are more exposed than in the minor groove. These interactions are dependent on electrostatic interactions and hydrogen bonds. The particular nature of each nucleotide allows the formation of specific hydrogen bonds between the functional groups of the nitrogen bases pointing outwards of the helix with the aminoacids of the protein [Alberts et al., 2002].

It is described that specific DNA-binding proteins also can bind DNA non-specifically. This non-specific binding facilitates the scanning of the recognition sequence by diffusion of the protein along the DNA. However, non-specific binding is typically several orders of magnitude weaker –as reported by their dissociation constants– than specific binding [Voet et al., 2013; Cox and Nelson, 2008]. Some structural motifs are characteristic of the specific recognition of DNA sequences, and these are exposed in the following lines.

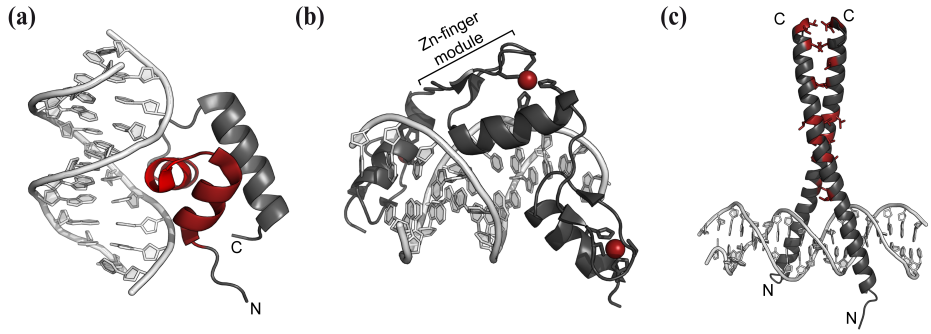

Figure 2.8: Specific DNA-binding motifs. Structure of the DNA-binding domain of the Lac repressor (NMR structure, PDB entry 1LCD). The helix-turn-helix domain is highlighted in red, with the recognition helix in lighter red. b) Zn-finger DNA-binding domain. Three connected Zn-finger modules interacting with the DNA. The protein is drawn as a gray cartoon. The His and Cys residues are represented as sticks and the coordinated Zn^{2+} ions as red spheres (X-Ray structure, PDB 1A1L). c) Leucine zippers motif. Two identical α helices dimerize by means of hydrophobic interactions between the side chains of leucines and valines, represented as red sticks (X-Ray structure, PDB 1YSA). Figures arranged with PyMOL (Schrödinger, LLC).

Helix-turn helix

The helix-turn-helix (HTH) domain is a $\sim$20 aminoacids structure consisting on two adjacent α helices separated by a short (3-4 aminoacids) turn, forming an angle of $\sim$120 between both helices [Voet et al., 2013]. One of the α helices, known as *recognition helix*, establishes specific hydrogen bonds with the base pairs of the DNA located on the major groove, and the interaction is further stabilized by additional contacts with the phosphate backbone. The second helix of the domain is not implicated on sequence discrimination, and just contributes to the stabilization of the interaction. The structural arrangement of both helices is conserved between them but the recognition of the DNA can occur in different geometrical orientations [Bloomfield et al., 1999]. Usually these domains work as symmetric homodimers with the recognition helices separated by one turn of the DNA helix (3.4 nm) and, as a result, they recognize palindromic sequences. By way of example, restriction enzymes, proteins that cleave the DNA at well-defined palindromic sites, recognize its binding sequence in a HTH-dependent fashion.

It is not uncommon that the binding of HTH domains results in distortion of the B-DNA structure. A well-known example is the Lac repressor, a protein regulating the synthesis of lactose in *E. coli* [Rohs et al., 2010]. The protein binds as a dimer to the DNA and upon recognition of the specific sequence, a structural change in the protein kinks the DNA by disrupting a stacking interaction between two adjacent bases (Figure 2.8a).

Zinc fingers

This motif was originally described as exclusively present in eukaryotes, where is the most widespread sequence-specific DNA-binding domain. Today Zinc-fingers are known to be present also in prokaryotes, and its structure resembles the helix-turn-helix motif, with an α helix as the main DNA-binding motif.

Zn-fingers are small DNA binding motifs of $\sim$30 aminoacids typically folded in two adjacent β sheets followed by an α helix [Voet et al., 2013; Sinden, 1994]. The major characteristic of their structure is the tetrahedral coordination of a Zn^{2+} ion by two conserved cysteines located at one of the β sheets and two histidines near the carboxy terminus of the α helix (Figure 2.8b). There are alternative residues coordinating the Zn ion, and different submotifs can be classified attending to the amminoacids involved. The coordination of the Zn ion keeps the α helix and the β sheets together and folded [Bloomfield et al., 1999]. The interaction with the DNA occurs between aminoacids of α helix domain and 2-3 base pairs of DNA. The α helix provides additional non-specific contacts with the phosphate backbone. In eukaryotes Zn-fingers do not interact in its basic modular form but forming tandem repeats twisting around the DNA [Watson et al., 2014]. This permits the recognition of long non-symmetric DNA sequences located at both the major and the minor grooves of the DNA. For instance the chromosome regulator protein CTCF involves 11 Zn-fingers to recognize a 14 bp sequence [Kim et al., 2007]. In contrast, prokaryotic Zn fingers can be found as a single module establishing additional interactions with the DNA not involving the β sheets nor the α helix [Malgieri et al., 2015].

Leucine zippers

Leucine zippers can be found in both prokaryotes and eukaryotes but are mainly characteristic motifs of eukaryotic transcription factors.

Leucine zippers are a family of DNA-binding domains resulting from coiled-coil dimerization of two α helices. The formation of the dimer is dependent on the interaction between a band of regularly-spaced hydrophobic residues [Voet et al., 2013]. The hydrophobic aminoacids responsible for dimerization are leucines (or valines) where the side chains of one helix interact with the side chains of the complementary helix. Therefore, leucine zippers are dimerization domains and not strictly speaking DNA binding domains. Nonetheless, dimerization results in the formation of an active DNA binding domain located at the amino-terminus (Figure 2.8c). The DNA binding domain is enriched in basic residues, where each monomer interacts with opposite sides on the major groove of the DNA in a tweezers fashion.

References

Adams, R. A. *Cálculo.* Pearson - Addison Wesley, 6th edition, 2009.

Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., and Walter, P. *Molecular Biology of the Cell.* Garland Science, 4th edition, 2002.

Bloomfield, V. A., Crothers, D. M., and Tinoco, I. J. *Nucleic acids: Structure, properties, and functions.* University Science Books, 1999.

Bosco, A., Camunas-Soler, J., and Ritort, F. Elastic properties and secondary structure formation of single-stranded DNA at monovalent and divalent salt conditions. *Nucleic Acids Res.*, 42(3): 2064–74, 2014.

Calladine, C. R. and Drew, H. R. *Understanding DNA. The molecule and how it works.* Academic Press, 2nd edition, 1997.

Cech, T. R. Crawling out of the RNA world. *Cell*, 136(4):599–602, 2009.

Cox, M. M. and Nelson, D. L. *Lehninger. Principles of Biochemistry.* Palgrave Macmillan, 4th edition, 2008.

Dessinges, M. N., Maier, B., Zhang, Y., Peliti, M., Bensimon, D., and Croquette, V. Stretching single stranded DNA, a model polyelectrolyte. *Macromolecules*, 89(24):248102, 2002.

Fuller, F. B. The writhing number of a space curve. *Proc. Nat. Acad. Sci. USA*, 68(4):815–9, 1971.

Griffiths, A. J. F., Miller, J. H., Suzuki, D. T., Lewontin, R. C., and Gelbart, W. M. *An Introduction to Genetic Analysis.* Pearson, 7th edition, 2002.

Hormeno, S., Moreno-Herrero, F., Ibarra, B., Carrascosa, J. L., Valpuesta, J. M., and Arias-Gonzalez, J. R. Condensation prevails over B-A transition in the structure of DNA at low humidity. *Biophys. J.*, 100(8):2006–15, 2011.

Kim, T. H., Abdullaev, Z. K., Smith, A. D., Ching, K. A., Loukinov, D. I., Green, R. D., Zhang, M. Q., Lobanenkov, V. V., and Ren, B. Analysis of the vertebrate insulator protein CTCF-binding sites in the human genome. *Cell*, 128(6): 1231–45, 2007.

Klug, W. S., Cummings, M. R., and Spencer, C. A. *Concepts of Genetics.* Pearson, 8th edition, 2005.

Kornberg, A. and Baker, T. A. *DNA Replication.* W. H. Freeman, 2th edition, 1992.

Lazcano, A. and Miller, S. L. The origin and early evolution of life: prebiotic chemistry, the pre-RNA world, and time. *Cell*, 85(6):763–8, 1996.

Malgieri, G., Palmieri, M., Russo, L., Fattorusso, R., Pedone, P. V., and Isernia, C. The prokaryotic zinc-finger: structure, function and comparison with the eukaryotic counterpart. *FEBS J.*, 282(23): 4480–96, 2015.

Marin-Gonzalez, A., Vilhena, J. G., Perez, R., and Moreno-Herrero, F. Understanding the mechanical response of double-stranded DNA and RNA under constant stretching forces using all-

atom molecular dynamics. *Proc. Natl. Acad. Sci. U.S.A.*, 114(27):7049–54, 2017.

Marko, J. F. and Siggia, E. D. Stretching DNA. *Macromolecules*, 28(26):8759Ű70, 1995.

Mendel, J. G. Versuche über plflanzenhybriden. *Verhandlungen des naturforschenden Vereines in Brünn*, 1866.

Moroz, J. D. and Nelson, P. Torsional directed walks, entropic elasticity and DNA twist stiffness. *Proc. Nat. Acad. Sci. USA.*, 94(26):14418Ű22, 1997.

Nelson, P. *Biological physics: Energy, information, life.* WH Freeman, 9th edition, 2007.

Odijk, T. Stiff chains and filaments under tension. *Annu. Rev. Biochem.*, 28: 7016–18, 1995.

Phillips, R., Kondev, J., and Theriot, J. *Physical Biology of the Cell.* Garland Science, 1st edition, 2009.

Rohs, R., Jin, X., West, S. M., Joshi, R., Honig, B., and Mann, R. Origins of specificity in protein- DNA recognition. *Annu. Rev. Biochem.*, 79:233–69, 2010.

Saenger, W. *Principles of Nucleic acids structure.* Springer, 2nd edition, 1988.

Sinden, R. R. *DNA structure and function.* Academic Press, 1st edition, 1994.

Strick, T., Allemand, J., Croquette, V., and Bensimon, D. Twisting and stretching single DNA molecules. *Prog. Biophys. Mol. Biol.*, 74(1-2):115–40, 2000.

Voet, D., Voet, J., and C., P. *Principles of Biochemistry.* John Wiley and Sons, 4th edition, 2013.

Wang, J. C. DNA supercoiling and its effects on the structure of DNA. *J. Cell. Sci. Suppl.*, 1:21–9, 1984.

Watson, J. D., Baker, T. A., Bell, S. P., Gann, A., Levine, M., and Losick, R. *Molecular Biology of the gene.* Pearson, 7th edition, 2014.

White, J. H. Self-linking and the Gauss integral in higher dimensions. *Am. J. Math.*, 91:693–728., 1969.

3

Magnetic tweezers

Chapter overview. Magnetic tweezers were developed less than twenty years ago. Despite their novelty, they have become one of the most widely used techniques for the manipulation and study of single molecules. This is a consequence of their simple implementation compared with other single-molecule techniques, their ability to easily apply torque and their power to obtain parallel measurements. In this chapter the technique will be explained in detail, including experimental examples and simple modifications to improve the capabilities of the system.

3.1 Introduction

Magnetic tweezers (MT) are a single molecule technique that allows applying forces and torque on biopolymers. This technique is based on the application of a forcs on magnetic beads as a result of the approximation of a magnetic field. The bead is tethered to a surface through the molecule of interest, which is extended upon the application of force. Accordingly, MT do not entrap the bead as in optical tweezers but pull from it, being incorrectly coined magnetic *tweezers* due to other similarities with its optical counterpart. A preliminary MT design is attributed to Steve Smith from the group of Prof. Carlos Bustamante [Smith et al., 1992]. In this implementation, an inverted microscope was used to image beads under the combined action of a drag force and a magnetic force, both parallel to the surface. In their set up, the measurement of the force was significantly affected by the interaction of the bead with the surface and exerting torques was not possible. Nonetheless, the authors already suggested the idea of rotating the magnetic field to apply torques to supercoil DNA molecules. MT as are known today were developed in 1996 by the group of Prof. Vincent Croquette [Strick et al., 1996] from the University Paris VII. In their configuration, magnets are located above the surface of the microfluidic cell and since the bead is stretched vertically, surface effects are negligible. This allows for a straightforward rotation of the magnetic field by rotating the magnets above the sample, exerting a torque on the tethered molecules. The determination of molecular extension in the z axis is possible thanks to the implementation of an algorithm based on the analysis of the diffraction rings of the beads.

The MT technique allows to perform experiments at constant force and torque without requiring feedback operation. The wide range of forces that can be applied

with MT –0.01 pN to 50 pN– have positioned them as a powerful force-spectroscopy technique to characterize the mechanical response of biopolymers, particularly nucleic acids. The great advantage of applying torque set MT as an invaluable tool to study the topology of nucleic acids as well as its regulation by proteins [Koster et al., 2005]. In addition, the ability to efficiently obtain parallel real-time measurements of single-molecules, has increased the rate of data acquisition, allowing the access to elusive features of protein-dependent DNA regulation with unprecedented resolution [De Vlaminck et al., 2011; Berghuis et al., 2011]. This chapter presents a complete and didactic description of the technique at both hardware and software levels, as well as guidelines for precise operation. The second part of the chapter presents relatively simple improvements of the technique, including several examples of application to the study of DNA-protein interactions.

3.2 Experimental set up

MT are a single-molecule technique whose construction and design are relatively straightforward compared with other single-molecule techniques. In essence, a MT setup is an inverted microscope designed to precisely control the focus to visualize micrometer beads on a fluidics cell (Figure 3.1). The employed beads are fabricated of a polymeric matrix with inclusions of ferrite nanoparticles, and consequently they are superparamagnetic. Rare earth permanent magnets are located above the microfluidic chamber, and their position is controlled with a motorized stage to modify the influence of the magnetic field on the beads. For paramagnetic beads, the action of an external magnetic field $\vec{B}$ induces a global magnetic moment on the bead $\vec{\mu}$, and the force applied $\vec{F}$ can be written as

$$\vec{F} = \frac{1}{2}\nabla(\vec{\mu}(\vec{B}) \cdot \vec{B}). \tag{3.1}$$

For a given position of the magnets, the exerted force is constant since the scale of the gradient of the magnetic field (mm) is much larger than the scale of the thermal fluctuations of the bead (from tens to hundred of nm). Additionally, if the magnets are properly aligned with the bead, the gradient of magnetic field in the xy plane is negligible compared with the gradient in the z axis, and thus the contribution of lateral forces can be omitted. The upward force exerted on the bead stretches a biopolymer –in the most common case DNA– which tethers the bead to the surface of the microfluidic chamber.

The most common procedure to prepare DNA molecules for MT experiments is by PCR (polymerase chain reaction). The DNA substrates are labeled with biotins and digoxigenins at each end of the molecule to permit the interaction with the bead and the surface, respectively. For that purpose, the surface is functionalized with anti-digoxigenin antibodies (anti-DIG) and superparamagnetic streptavidin-coated beads establish a strong bond with biotins and are commercially available. This functionalization procedure is sufficient for forces < 50 pN, within the range of most biological processes. The pulling force stretches the molecule, lifting the bead above the surface, and as a consequence the distance of the bead to the objective focal plane (OFP)

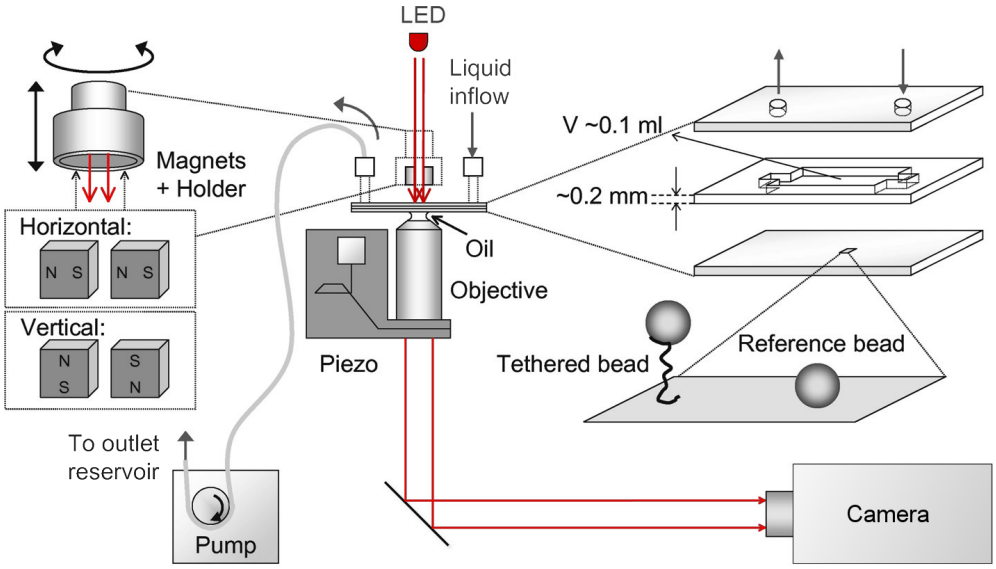

Figure 3.1: Components of a magnetic tweezers (MT) setup. MT is based on inverted microscope with piezo-controlled focus and designed to image micron-sized beads. The sample is illuminated with LED light and images are captured with a video camera. Beads are anchored to the surface of a microfluidic chamber through the substrate of interest. Beads and molecules are injected on the microfluidics chamber with a specially designed adapter to facilitate liquid exchange with a syringe pump. The magnets that can be translated vertically and rotated to apply force and torque to the molecules are located above the microfluidics cell. Adapted from Vilfan et al. [2009].

is modified. Images of beads are captured by means of videomicroscopy and computationally analyzed to determine the extension of the molecule from the profile of diffraction rings. The profile of rings is characteristic of the distance of the bead to the OFP.

The major advantage of MT with respect to other single-molecule techniques relies in their ability to apply torque. A particle with magnetic moment $\vec{\mu}$ in a magnetic field $\vec{B}$ experiences a torque $\vec{\tau}$ given by [Tipler and Mosca, 2010]

$$\vec{\tau} = \vec{\mu} \times \vec{B} \tag{3.2}$$

that tends to align both vectors. For paramagnetic beads, the magnetic moment is an induced magnetic moment $\vec{\mu} = \vec{\mu}(\vec{B})$, where a perfect alignment between $\vec{\mu}$ and $\vec{B}$ is expected. Nonetheless, the magnetic susceptibility of the beads is not isotropic and as a consequence there is a component of $\vec{\mu}$ not aligned with $\vec{B}$ [van Oene et al., 2015]. Since the rotation of the magnets changes the direction of the magnetic field, a torque is generated following equation 3.2, resulting in a coupled rotation of the bead and the magnets, a phenomenon that will not occur if $\vec{\mu}(\vec{B})||\vec{B}$. The exerted torque on the bead is engaged to the tethered molecule and by measuring the extension in response to the applied turns, it is possible to characterize the mechanical response of nucleic acids to the applied torsion. Examples of the torsional properties of nucleic

acids as well as its implication on protein activity are presented at the end of this chapter and in following chapters.

3.2.1 Magnetic Tweezers setup used in this thesis

The setup described here is similar in design to that designed by Seidel et al. [2004]. The system was assembled and modified by Fernando Moreno-Herrero, PhD and the updating of the software was a contribution by Fernando Moreno-Herrero, PhD and Maria Teresa Arranz, M. Eng.

Optics
Micrometer-sized beads are visualized with a 100X oil-immersion objective (Olympus ULSAPO 100XO; numerical aperture = 1.2). The focal plane can be adjusted by vertically moving the objective using a micrometer stage (M-UMR5.25, Newport). Precise focusing is achieved by controlling the axial position of the objective with a piezo electric stage (P-726, Pyhsik Instrumente, Germany). The sample is illuminated using Köhler illumination with a high power LED emitting at 635 ± 15 nm (LED635L, Thorlabs, USA) as light source. The outcoming light from the objective is redirected using a 2 inches mirror at 45° (BB2-E02, Thorlabs, USA). The reflected image is focused onto the camera's chip using a planoconvex lens with a focal distance of 200 mm (PAC064AR.14, Newport, USA) mounted on a lens tube.

Motors and Magnets
Translation and rotation of the magnets are carried out using stepper motors. A linear motor (M-126.PD1, Pyhsik Instrumente, Germany) is mounted on a vertical z-axis bracket (M-125.90, Pyhsik Instrumente, Germany) and acts as translational device. An aluminium block screwed to the linear motor stage supports the rotation motor and a cylindrical magnet holder fixed to a bearing. The rotation of the motor is transmitted to the magnet holder using aluminium gears and a Synchroflex belt (RS Components). Gold coated neodymium (NdFeB) cubic magnets ($5\times5\times5$ mm, W-05-N50-G, Supermagnete GmbH, Germany) are mounted on an aluminium or sweet iron head depending on the desired final alignment. The alignment of the magnets with the optical axis is controlled with two micrometer stages (M-SDS65 Newport, USA) below the translation motor bracket.

Camera.
Video images of the beads are captured with a CMOS camera (EoSens CL, Mikrotron GmbH, Germany) with maximum acquisition frequency of 506 Hz at full-resolution and 120,000 Hz at reduced resolution.

Software
The full system is controlled with a custom-written LabVIEW software (National Instruments, USA) in its version 11. The software allows real-time analysis of the spatial coordiantes of up to 30 beads at 120 Hz with a lateral spatial resolution of 2 nm.

Microfluidics
Flow cells are fabricated using two coverslips (BB024060A1, Menzel Gläser, Germany) cleaned by sonication for 15 minutes in isopropanol. Surface coverslips are coated with 3 µl of a dilution 1:30 of magnetic beads (Dynabeads MyOne, Invitrogen) in ethanol and gently sparse with the side of a pipette tip before functionalization with a 1% solution of polystyrene in toluene. This procedure ensures the presence of fixed reference beads necessary for accurate measurements of extensions. Inlet holes in the upper coverslips are drilled using a laser engraver (VLS2.30, Universal Laser System, USA). Flow cells are assembled with one or two layers of parafilm (Parafilm M, USA) between the upper and the surface coverslip, leaving a central channel of 0.2×6×2.5 mm or 0.1×6×2.5 mm for double or single parafilm layers, respectively. Parafilm and coverslips are sealed by incubation on a heating block. Final functionalization of the cell is achieved by incubating anti-digoxigenin antibodies (11333089001, Roche, Switzerland) overnight at 4°C or for 3 hours at 37°C. Flow cells are mounted onto the microscope by an alumnium plate screwed to a polyether ether ketone (PEEK) cage whose xy position can be controlled with two linear translation stages (M-UMR5.25, Newport, USA) to select the observation area. The PEEK cage is connected with silicone tubing to a syringe pump (NE-1000, New Era Pump Systems, USA) to facilitate buffer exchange and to control the flow rate. Instantaneous starting/stopping of the flow is achieved with a solenoidal valve (LHDA0531115H, Lee Company, USA).

3.3 The DNA-bead system and force calibration

At a constant applied force F, the DNA-bead system can be considered as an inverted pendulum of extension l (Figure 3.2). Brownian fluctuations drive the bead our from the equilibrium position a distance δx with an angle θ with respect the force axis. In an out of equilibrium position, the bead experiences a restoring force F_r which leads the system back to the equilibrium position. This force is given by $F_r = F \sin\theta$. Considering the system as a harmonic oscillator, the restoring force can be also written as $F_r = k\delta x$, where k is the spring constant, known in this context as magnetic *trap* stiffness. According to the energy equipartition theorem, the average potential energy of this harmonic system equals the average kinetic energy of the environment, one-half $k_B T$ per degree of freedom:

$$\frac{1}{2} k \left\langle \delta x^2 \right\rangle = \frac{1}{2} k_B T. \tag{3.3}$$

and thus the spring constant reads as $k = k_B T / \left\langle \delta x^2 \right\rangle$. For small angular displacements the change in extension is negligible ($\delta z \ll \delta x$) and then $\theta = \delta x / l$. Given previous assumption, it is possible to apply the small-angle approximation ($\sin\theta \approx \theta$) to the restoring force and then,

$$F_r = F \frac{\delta x}{l} \tag{3.4}$$

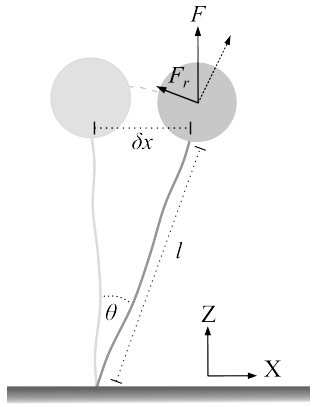

Figure 3.2: Geometry of the DNA-bead system. The applied magnetic force F stretches the molecule to an extension l. Brownian fluctuations displace the molecule from its equilibrium position a distance δx. This results in a tangential restoring force F_r, which can be approximated to be linerly dependent on δx.

where clearly can be identified the value of the spring constant, $k = F/l$. Finally, this result can be combined with the result obtained for the spring constant from equation 3.3 to obtain

$$F = \frac{k_B T l}{\langle \delta x^2 \rangle}, \tag{3.5}$$

which is the relation typically used in magnetic tweezers experiment to determine the applied force. Notice, however, that the force is determined from equilibrium measurements of the fluctuations and the extension. Therefore, it is not possible to obtain real-time measurements of the applied force as in optical tweezers or AFM. The procedures to extract the extension of the molecule l and the transversal fluctuation $\langle \delta x^2 \rangle$ are described in the following section.

3.4 Determination of bead's coordinates

In contrast to other single-molecule techniques, in MT the positions of the bead are fully determined from image analysis. Importantly, MT are not strictly speaking force sensors since the determination of the applied forces relies on sampling the transversal fluctuations of the bead at the equilibrium for a fixed position of the magnets. Thence, precise determination of bead coordinates is not only a requirement for measuring molecular extensions but in addition for the characterization of applied forces. Typically, the analysis of the coordinates is performed in *real-time* allowing the experimentalist to take decisions based on the observation. This is an extraordinary advantage for the user with a huge computational cost, as the coordinates of several beads have to be extracted in the millisecond time-scale. From the previous arguments, it is clear that highly optimized algorithms are necessary to obtain fast and precise measurements of the spatial coordinates of the bead.

3.4.1 Analysis of in-plane positions

The procedure here described is based on the ideas introduced by Gosse and Croquette [2002] and Gelles et al. [1988]. In order to measure the horizontal positions of the bead with sub-pixel accuracy, an average profile of the intensities $I(x)$ is obtained from the video image taking 5 rows (for the x axis) or columns (for the y axis) around the center of a region of interest (ROI). Thermal fluctuations drive the bead out of the central position of the ROI, resulting in a new profile displaced with respect to the symmetry axis by a distance δx, and the new profile reads $I(x+\delta x)$. To detemine the parameter δx, the profile $I(x+\delta x)$ is compared with its specular image, $I(-x-\delta x)$ (Figure 3.3a) via cross-correlation, $C(\tau) = I(x+\delta x)I(-x-\delta x)(\tau)$, where τ states for the displacement between both functions (Figure 3.3b). It is intuitive that the maximum cross-correlation occurs at $\tau = 2\delta x$. Because of its improved computational efficiency, the cross-correlation function is calculated in Fourier space using the Fast Fourier Transform algorithm. The *Correlation theorem* relates the correlation function with the Fourier transform ($\mathcal{F}\{...\}$) of two functions [Chatfield, 2013; Press et al., 2009] and hence the correlation function of both profiles can be obtained in the frequency domain as

$$C(\tau) = \mathcal{F}^{-1}\{\mathcal{F}\{I(x+\delta x)\} \cdot \mathcal{F}^{*}\{I(-x-\delta x)\}\}, \tag{3.6}$$

where $\cdot$ states for the the pointwise-product (elementwise-product) of the Fourier coeficients of each function, * denotes the complex conjugate, and $\mathcal{F}^{-1}\{...\}$ indicates the inverse Fourier transform. The value of τ maximizing the correlation function can be obtained by fitting a parabola to the $\pm$ 2 or $\pm$ 3 points around the maximum correlation parameter [van Loenhout et al., 2012a]. Finally, the distance to the symmetry axis can be obtained from $^1/_{2\tau}$ and the ratio µm/pixel. According to Gosse and Croquette [2002], this algorithm leads to a sub-pixel resolution of $^1/_{10}$ to $^1/_{100}$ of pixel. We achieve 1-2 nm of spatial resolution in-plane with the MT setup used in this work.

3.4.2 Analysis of axial positions

The procedure to obtain the distance between the bead and the surface –which is comparable to molecular extension, l– relies on the analysis of the diffraction rings of the bead described by Gosse and Croquette [2002]. Using nearly-monochromatic light to illuminate the sample results in the formation of an Airy pattern around the bead. The profile of rings is highly dependent on the distance to the objective focal plane (OFP) and this is going to be exploited to determine l. The first step involves the generation of a calibration matrix, from now on Lookup Table (LUT), composed by intensity profiles of the rings of a reference bead at different focal planes (Figure 3.4a). The LUT is accurately extracted by translating the objective position using a piezoelectric stage. In order to determine the distance to the OFP f of a certain bead, the average intensity profile I of the bead is obtained during the tracking routine (Figure 3.4b) and compared to the LUT by means of the χ^2 function (Figure 3.4c).

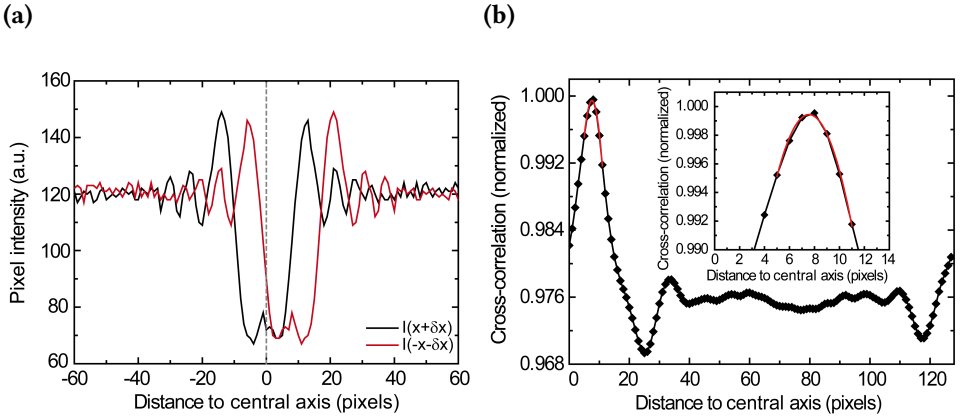

Figure 3.3: Determination of in-plane coordinates. a) x-axis profile of a 4 pixels (504 nm) artificially displaced bead with respect to the central axis (black). In red, mirror profile of the same image. b) Circular cross-correlation of the profiles in panel a. Inset: detail of the maximum correlation region and parabola fit showing the maximum correlation around $\tau =$ 8 pixels indicating a displacement of 4 pixels.

For a given distance f displaced by the objective in the LUT, the χ^2 function reads:

$$\chi^2(f) = \sum_{i=1}^{N} \left[I(i) - \mathrm{LUT}(f, i) \right]^2 , \tag{3.7}$$

where i runs the position of the profile in a N pixels ROI. The precise determination of the value of f minimizing χ^2 can be found by fitting a parabola to the $\pm$ 2 or $\pm$ 3 points around the minimum value of χ^2 (Figure 3.4c). This procedure is a computationally efficient way to find the distance to the OFP of the tracked bead with a resolution of a few nanometers, without the necessity of exploring a large LUT with different focal planes.

Nevertheless, the change of focus inside the liquid chamber is different to the distance traveled by the objective during LUT acquisition because of the different refraction indexes of the water in the experimental buffer and the glass of the microfluidic cell [Hell et al., 1993]. Accounting for the refraction occurring at the interface, it is possible to obtain a corrected change of focus $\tilde{f}$ as

$$\tilde{f} \simeq f \left(\frac{n_{water}}{n_{glass}} \right) \tag{3.8}$$

where n refers to the refractive index of each media. For the case of a water-glass interface this correction factor equals 0.88. Notice that this is an approximation formula which is not accurate for high numerical aperture objectives [Hell et al., 1993; van Loenhout et al., 2012a]. Once calculated the corrected distance to the OFP $\tilde{f}$, the extension of the molecule l is obtained by subtracting to this the corrected distance

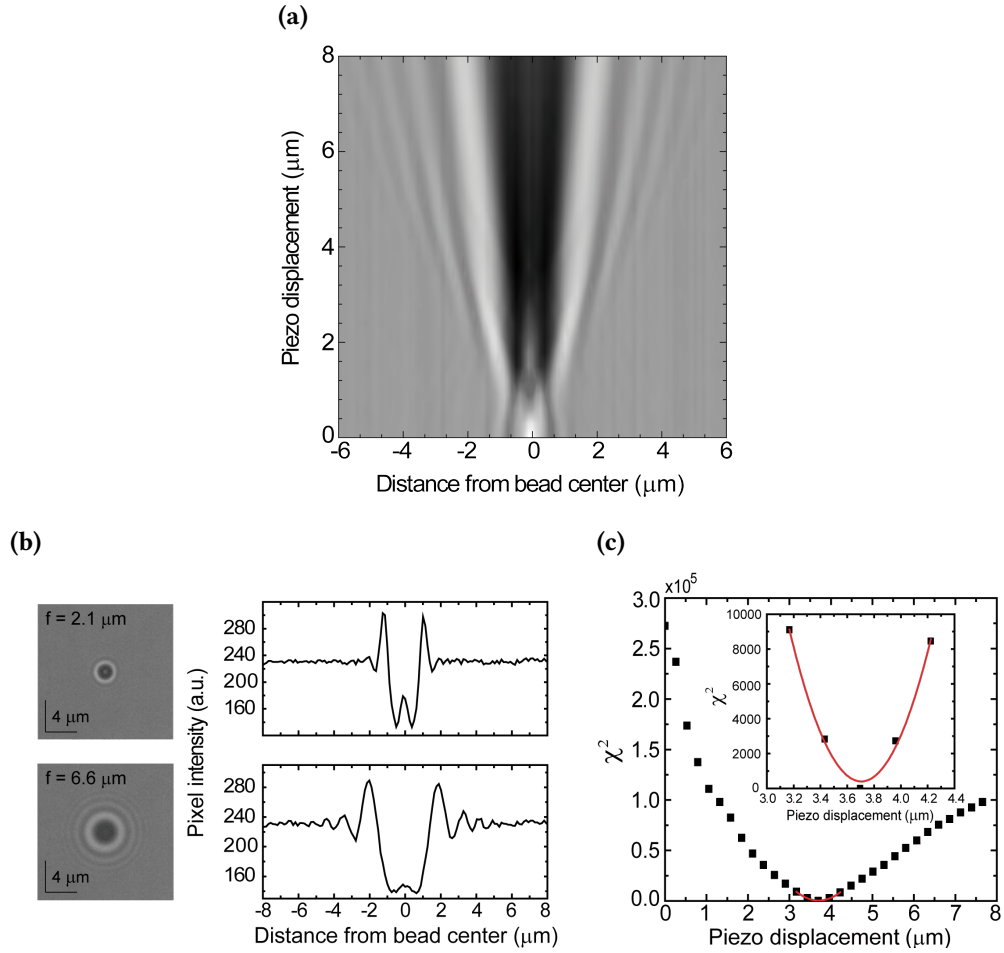

Figure 3.4: Determination of axial positions in MT. a) Calibration pattern (LUT) for profile intensities at different objective's focal planes (OFP) obtained in 300 nm steps for a total traveled distance of 8 μm. b) Examples of the diffraction rings and radial profiles of beads at 2.1 μm and 6.1 μm below the OFP. c) χ^2 function for a bead located 3.6 μm to the objective focal plane. Inset: Points around the minimum difference to the calibration profile and parabola fit.

to the OFP of a reference bead fixed to the surface of the cell: $\tilde{f}_{\text{Ref}}$, as $l = \tilde{f} - \tilde{f}_{\text{Ref}}$ (Figure 3.5a). This correction ensures proper determination of molecular extension l independently of the arbitrary selection of the focus during the acquisition of the LUT. In addition, this correction overcomes drift effects on the vertical axis arising due to changes of temperature or mechanical perturbations.

This method for the determination of molecular extension l is only correct if the molecule is anchored at the south pole of the bead. In a more realistic scenario, the DNA is anchored in a position below the equator with an angle α with respect to the south pole (Figure 3.5b). However, the procedure previously exposed, does not consider the fraction of the molecular extension hidden by the bead (l_b) and then

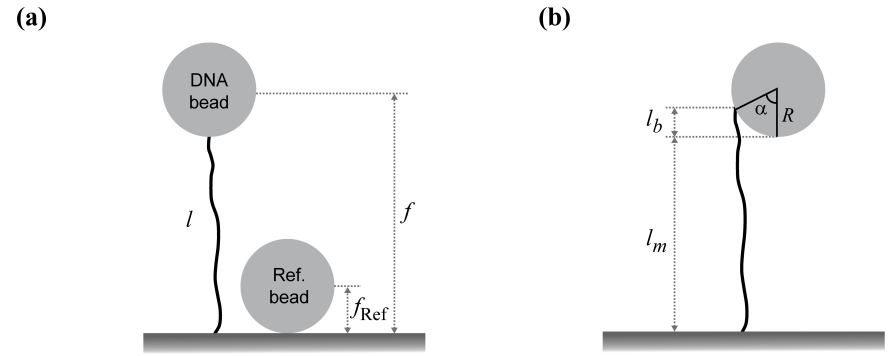

Figure 3.5: Determination of molecular extension from focal distances. a) Molecular extension l is determined from the difference in the focal distances of the DNA bead f with a reference bead f_{Ref}. b) Realistic situation where the DNA is anchored with certain angle α to the south pole of the bead. The measured extension l_m underrates the real molecular extension l by a quantity l_b.

the real extension is $l = l_m + l_b$, where l_m is the measured extension. The value of l_b can be determined attending to the geometry of the bead as $l_b = R(1 - \cos\alpha)$, where R is the radius of the bead. The parameter α can be obtained by rotating the magnets and following the procedure that will be described later in section 3.7.2. The details are skipped here since its contribution depends on the relation between the size of the bead and the length of the tether, being this correction ommited in routine measurements.

The spatial resolution obtained with this procedures depends on several factors, as magnification or the number of pixel of the camera. In the setup used here, we determined a spatial resolution in the vertical axis of 5 nm for surface-stuck beads.

3.5 Calculation of forces and bead's equation of motion

Accurate determination of applied forces relies on accurate determination of the variance of in-plane fluctuations $\langle \delta x^2 \rangle$. The calculation of the variance could be performed in real space following the well-known formula

$$\langle \delta x^2 \rangle = \frac{1}{N} \sum_{i=1}^{N} (x_i - \mu_x)^2 \tag{3.9}$$

where N is the number of data points in the set and $\mu_x = 0$ is the mean distance of the bead to the center of the x axis. However, this formulation is impractical. First, this approach is very sensitive to the mean value and thus sample drift or outlier values result in a dramatic difference between the variance estimator and the true variance. Second, this approximation does not take into account the stroboscopic effects arising from the finite acquisition frequency of the camera. The limited acquisition frequency of the camera leads to an inaccurate sampling of bead's motion and consequently to an incorrect determination of the variance as will be expanded in the following section. Both problems can be overcome by extracting the mean

square displacement $\langle \delta x^2 \rangle$ in Fourier space. This procedure takes advantage of the *Parseval's theorem* [Press et al., 2009; Chatfield, 2013] which states that the integral of the modulus of a square function equals the integral of the square modulus of the Fourier transform. The square modulus of the real and imaginary parts of the Fourier transform is defined as power spectral density (PSD, $P_x(f)$):

$$P_x(f) \equiv |\mathcal{F}\{x(t)\}|^2 = \mathcal{F}\{x(t)\} \cdot \mathcal{F}^*\{x(t)\}, \tag{3.10}$$

and then the variance can be obtained from Parseval's relation as

$$\langle \delta x^2 \rangle = \int_0^\infty P_x(f)\, df \tag{3.11}$$

This expresion can be combined with equation 3.5 to determine the applied force. Remarkably, this approximation for calculating the variance based on the fluctuation analysis in the frequency domain allows to correct the PSD to prevent camera-derived artifacts. The details of these procedures will be explained in section 3.5.2.

3.5.1 Equation of motion

In the previous section it has been exposed the methodology followed to determine forces based on the PSD of the in plane fluctuations. Here, the theoretical PSD of the bead is going to be derived from the equation of motion of a bead in a harmonic potential. This can be described by the following Langevin equation [Gosse and Croquette, 2002]:

$$m\frac{d^2x(t)}{dt^2} = -\gamma\frac{dx(t)}{dt} - kx(t) + F_{\mathrm{L}}(t) \tag{3.12}$$

where the first term is the inertial force with m the mass of the bead. The term $kx(t)$ states for the action of a restoring force on the bead from a harmonic potential well with stiffness $k = F/l$. The Langevin force $F_{\mathrm{L}}(t)$ describes a random force on the bead due the collisions with the surrounding water molecules. The Langevin force equals $F_{\mathrm{L}}(t) = \sqrt{2\gamma k_B T}n(t)$ where $n(t)$ is the noise term, here a stochastic Gaussian process with zero-mean $\langle n(t) \rangle = 0$ and autocorrelation function $\langle n(t)n(t-t') \rangle = \delta(t')$ with δ the Dirac delta function. The first term at the right side of the equality is the drag force acting on the bead from Stokes' law where $\gamma = 6\pi\eta R$ is the hydrodynamic friction coefficient. Here η is the dynamic viscosity and R is the radius of the bead. At the typical acquisition frequencies used in MT experiments, the inertial component can be neglected[1] and the system can be considered as Brownian or overdamped

$$0 = -\gamma\frac{dx(t)}{dt} - kx(t) + F_{\mathrm{L}}(t). \tag{3.13}$$

[1]The characteristic inertial frequency $f_{inert} = \gamma/m \sim 17$ MHz from typical parameters $m = 10^{-15}$ kg, $R = 0.5{\cdot}10^{-6}$ m and $\eta = 10^{-3}$kg m^{-1}s^{-1} is much faster than most camera's acquisition frequencies. The assumption of a negligible inertial component is then justified.

Taking the Fourier transform of the previous equation,

$$0 = -\gamma\mathcal{F}\left\{\frac{dx(t)}{dt}\right\} - k\mathcal{F}\{x(t)\} + \mathcal{F}\{F_{\mathrm{L}}(t)\} \tag{3.14}$$

where $\mathcal{F}\{dx(t)/dt\} = -2\pi f i\mathcal{F}\{x(t)\}$ with i as the imaginary unit $i = \sqrt{-1}$. Solving for $\mathcal{F}\{x(t)\}$ results in

$$\mathcal{F}\{x(t)\} = \frac{\mathcal{F}\{F_{\mathrm{L}}(t)\}}{2\pi\gamma(f_c - fi)} \tag{3.15}$$

where it has been defined $f_c \equiv k/2\pi\gamma$. This is the corner or cut-off frequency of the system, whose inverse determines the relaxation time of the bead back to the equilibrium situation. The power spectrum can be obtained from its definition by taking the square of the absolute value of the previous equation, where the complex expression $|(f_c - fi)|^2 = f_c^2 + f^2$ and thus

$$P_x(f) = \frac{P_{F_{\mathrm{L}}}(f)}{4\pi^2\gamma^2}\frac{1}{(f_c^2 + f^2)}. \tag{3.16}$$

From the correlation theorem described in section 3.4.1 (equation 3.6) and the definition of PSD given in equation 3.10, it can be seen that the Fourier transform of the autocorrelation function equals the PSD [Press et al., 2009]. This special case of the correlation theorem is known as *Wiener-Khinchin theorem* and it can be applied to determine the two-sided PSD of the random force $P_{F_{\mathrm{L}}}$,

$$\begin{aligned} P_{F_{\mathrm{L}}} &= \int_{-\infty}^{+\infty} F_{\mathrm{L}}(t)F_{\mathrm{L}}(t-t')e^{-2\pi i f t'}dt' \\ &= 2\gamma k_B T\int_{-\infty}^{+\infty} n(t)n(t-t')e^{-2\pi i f t'}dt' \\ &= 2\gamma k_B T\int_{-\infty}^{+\infty} \delta(t')e^{-2\pi i f t'}dt' \end{aligned} \tag{3.17}$$

Since the Fourier transform of the Dirac delta, $\mathcal{F}\{\delta(x)\} = 1$,

$$P_{F_{\mathrm{L}}} = 2\gamma k_B T. \tag{3.18}$$

In this context there is no distinction between positive and negative frequencies, and, as a consequence, equation 3.18 is multiplied by a factor two to obtain the one-sided PSD [Press et al., 2009], and then $P_{F_{\mathrm{L}}} = 4\gamma k_B T$. The frequency-independent contribution of $F_{\mathrm{L}}(t)$ to every frequency of the PSD describes a contribution in the form of white noise. This result can be combined with equation 3.16 to derive an expression for the theoretical power spectrum as

$$P_x(f) = \frac{k_B T}{\pi^2\gamma}\frac{1}{(f_c^2 + f^2)}. \tag{3.19}$$

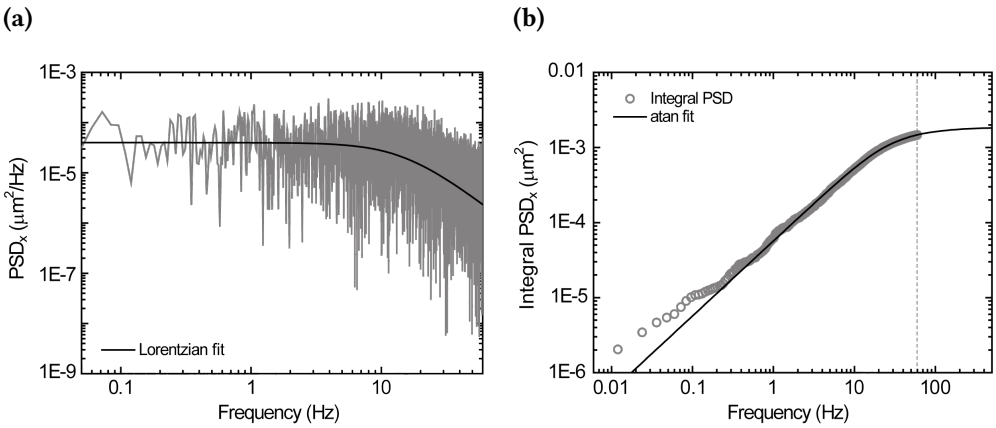

Figure 3.6: Force calibration in frequency domain. a) Experimental power spectral density (PSD) of the fluctuations on the x axis for a bead at $F \approx 5$ pN. Lorentzian fit results in a cut-off frequency $f_c = 15$ Hz. b) Integrated power spectrum and fit to equation 3.23, returning a variance of $\langle \delta x^2 \rangle = 20$ nm^2. The dashed line indicates the maximum sampling frequency.

This equation is a Lorentzian and can be used to fit the PSD to determine the value of f_c (Figure 3.6a). Importantly, this expresion can be integrated following Parserval's theorem (equation 3.11), and then an analytical result for the variance of the fluctuations can be deduced from the integral of the one-sided PSD,

$$\begin{aligned} \langle \delta x^2 \rangle &= \frac{k_B T}{\pi^2 \gamma} \int_0^\infty \frac{1}{f_c^2 + f^2} df \\ &= \frac{k_B T}{\pi^2 \gamma f_c^2} \int_0^\infty \frac{1}{1 + (f/f_c)^2} df, \end{aligned} \tag{3.20}$$

and by substitution the integral can be solved resulting in

$$\begin{aligned} \langle \delta x^2 \rangle &= \frac{k_B T}{\pi^2 \gamma f_c} \tan^{-1} \left(\frac{f}{f_c} \right) \Big|_{f=0}^{f=\infty} \\ &= \frac{k_B T}{2\pi \gamma f_c}. \end{aligned} \tag{3.21}$$

Taking the definition of f_c,

$$\langle \delta x^2 \rangle = \frac{k_B T}{k} \tag{3.22}$$

one recovers the result obtained in real space when the energy equipartition theorem is combined with the energy of harmonic potentials in one dimension. This is indeed a useful result which can be combined with equation 3.19 to obtain a fitting equation of the PSD,

$$\int_0^{f_{max}} P_x(f) df = \frac{2}{\pi} \langle \delta x^2 \rangle \tan^{-1} \left(\frac{f}{f_c} \right) \Big|_{f=0}^{f=f_{max}} \tag{3.23}$$

where the integral is evaluated numerically and both $\langle \delta x^2 \rangle$ and f_c are free parameters of the fit (Figure 3.6b). Then, f_c can be used to compute the radius of the bead R or the dynamic viscosity η if R is known. $\langle \delta x^2 \rangle$ can be directly plugged into equation 3.5 to determine the applied force. Importantly, the fit extrapolation to high frequencies allows overcoming to some extent the limitations of the acquisition frequency. Nonetheless, precise measurements involving high forces or short molecules require additional corrections and these are described in the following section.

3.5.2 Camera corrections

In the previous section it has been shown the methodology to extract the applied force based on the integration of the PSD. However, integration can only be performed over a finite number of frequencies due the limited sampling frequency of the camera, f_s. The *Nyquist-Shannon sampling theorem* states that the acquisition frequency must be at least twice the highest frequency of the system. Thus, for frequencies higher than the so-called Nyquist frequency, $f_{\text{Nyquist}} \equiv 1/2 f_s$ there is a loss of spectral information (Figure 3.7a). There are two major artifacts arising when the signal presents frequencies higher than f_{Nyquist}. First, the camera captures the image during a finite shutter or exposure time W, implicitly averaging the input image and consequently resulting in a blurred output image. In MT experiments this results in an apparent reduction of the variance (Figures 3.7a and b). Second, frequencies higher than f_s are included in the PSD at wrong frequencies. This artifact is known as back-folding or aliasing an it does not affect the total power but alters the shape of the PSD, understimating the corner frequency f_c. An everyday life example of these effects can be observed in the aparently static blades of rotating fans.

From the previously exposed arguments it is clear that proper characterization of the PSD is a must for accurate force calibrations. Therefore, corrections that prevent or at least dismiss artifacts arising from the limited acquisition frequency of the camera are desired. Here it is considered a theoretical description of motion tracking by cameras derived by Wong and Halvorsen [2006].

The measured motion of the bead $x_s(t)$ can be obtained by convolving a rectangular window with the real trajectory of the bead $x(t)$ (Figure 3.7b),

$$x_s(t) = \int x(t')H(t-t')dt \tag{3.24}$$

where $H(t)$ is a boxcar function defined as

$$H(t) = \begin{cases} \frac{1}{W} & 0 < t \leq W \\ 0 & \text{otherwise} \end{cases} \tag{3.25}$$

and acting as a moving average filter over the true motion $x(t)$. Then, it can be defined the sampled PSD $P_{x_s}(f)$ (Figure 3.7c) as

$$P_{x_s}(f) = P_x(f)P_H(f), \tag{3.26}$$

where $P_x(f)$ and $P_H(f)$ are the PSDs of the bead motion and the box function respectively. It is clear that the PSD corrected for motion blur can be obtained simply by dividing the measured PSD by the PSD of the rectangular function $P_{xs}(f)/P_H(f)$. The reader might notice that the Fourier transform of the rectangular function is the $\mathrm{sinc}(\pi fW)$ function and then the PSD is simply

$$P_H(f) = \left(\frac{\sin(f\pi W)}{f\pi W}\right)^2. \tag{3.27}$$

Combining this equation with the PSD for the true motion of the bead derived in the previous section (equation 3.19) and following equation 3.26

$$P_{x_s}(f) = \frac{k_B T}{\pi^2 \gamma} \frac{1}{(f_c^2 + f^2)} \left(\frac{\sin(f\pi W)}{f\pi W}\right)^2 \tag{3.28}$$

which can be integrated to yield the measured variance $\left\langle \delta x_s^2 \right\rangle$ (Figure 3.7d), and resulting in

$$\left\langle \delta x_s^2 \right\rangle = \frac{2k_B T\gamma}{k^2 W} \left[1 - \frac{1}{kW}\left(1 - e^{-\frac{kW}{\gamma}}\right)\right]. \tag{3.29}$$

Notice that the previous analysis does not take into account for aliasing effects. These are not going to affect the measured variance $\left\langle \delta x_s^2 \right\rangle$ and thus they are less critical than the motion blur correction in terms of force quantification. However, aliasing changes the shape of the PSD shifting the corner frequency of the system f_c to lower frequencies, which can be important if the experimentalist is interested in characterizing the magnetic trap k or its derived parameters. Considering that aliasing contributes to the power of a certain frequency by summing up the power of all the integer multiples n of the acquisition frequency f_s,

$$\begin{aligned} P_{x_s}^{alias}(f) &= \sum_{n=-\infty}^{+\infty} P_{x_s}(f + nf_s) \\ &= \sum_{n=-\infty}^{+\infty} \frac{k_B T}{\pi^2 \gamma} \frac{1}{(f_c^2 + (f + nf_s)^2)} \left(\frac{\sin(\pi W(f + nf_s))}{\pi W(f + nf_s)}\right)^2. \end{aligned} \tag{3.30}$$

This equation can be solved analytically for the case $f_s = 1/W$ [Lansdorp and Saleh, 2012] but in practice it is solved numerically considering a few terms of n, typically 1 and -1. Therefore, it is possible to obtain a PSD accounting for corrections to both aliasing and motion blur $P_x^{corr}(f)$ [te Velthuis A. J. W et al., 2010] as:

$$P_x^{corr}(f) = \frac{P_x(f) - P_{x_s}^{alias}(f)}{P_{x_s}}. \tag{3.31}$$

Albeit these corrections significantly improve the determination of the true PSD, if the frequency of the system is much higher than f_s, the user must find both instrumental and analytical alternatives to precisely characterize the applied forces.

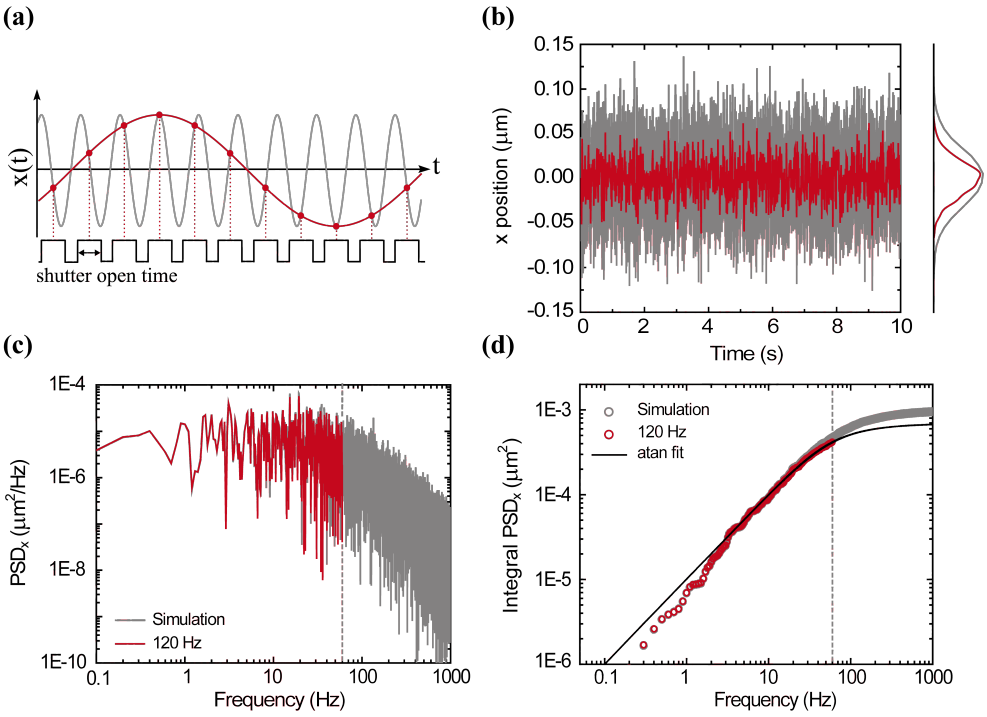

Figure 3.7: Effects of camera acquisition on the spectral analysis of bead's motion. a) Camera integration time artifact on the measured signal (red) of a sinusoidal wave (gray). Averaging during exposure time leads to a decrease of the measured frequency. b) Brownian dynamics simulation of the x axis motion of a 2 μm molecule at $F = 8$ pN with a 5 μs time step (grey) as described in te Velthuis A. J. W et al. [2010] and Beausang et al. [2007]. In red, bead motion after emulating camera acquisition at 120 Hz. c) Power spectral densities (PSD) of the simulated motion and after emulating camera acquisition. d) Integrated PSDs of simulated (gray symbols) and camera-emulated data (red symbols). Fitting to equation 3.23 (black line) underestimate the variance, resulting in a estimated force of ≈ 11 pN. Dashed vertical lines indicates the Nyquist frequency, f_{Nyquist}.

3.6 MT experiments under force and torque

The mechanical response of biopolymers can be characterized with MT. However, in contrast to optical tweezers or atomic force microscopy where once determined the stiffness of the trap/cantilever the force can be characterized in real-time, the determination of forces in MT relies on sampling the equilibrium extension of the molecule and the variance of the fluctuations for a fixed position of the magnets. The measurement time to characterize equilibrium extensions and variances depends on the applied force, ranging from a few seconds for forces > 4 pN to several minutes at low forces. An optimum measurement time can be selected attending to the Allan variance [Czerwinski et al., 2009; Lansdorp and Saleh, 2012] to account for sampling equilibrium parameters while noise derived from thermal or mechanical drifts is minimized. The interested reader is referred to Appendix B for an expanded

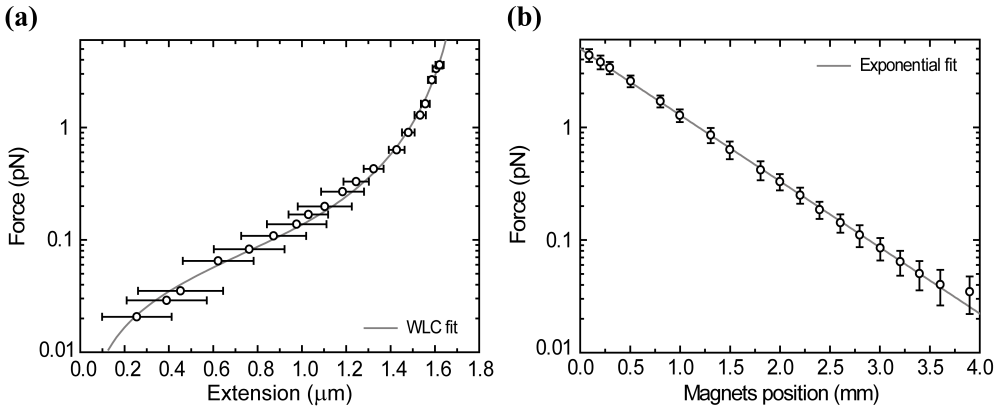

Figure 3.8: Calibration of Magnetic tweezers. a) Force-extension curve of a 6 kbp molecule and fit to the worm-like chain model (WLC). The output parameters from the fit are $\xi = 45$ nm and $L = 1.8$ μm in agreement with the reported value and with the expected contour length for a basepair distance of 0.34 nm. Error bars represent standard deviation (std) of the extension at a given force. b) Calibration curve for the applied force at different magnet positions relative to the upper surface of the microfluidic cell and exponential fit. Error bars represents the std.

explanation of the method. Next, by sequentially extracting molecular extension and force at different magnet positions it is possible to determine the mechanical parameters of biopolymers by fitting force-extension curves to the worm-like chain model (WLC, Figure 3.8a). The model depends on two parameters, i) the persistence length ξ which measures the bending stiffness, and ii) the contour length L which is simply the total length of the molecule. The model was described in detail on section 2.4.2. Remarkably, both parameters are well-characterized for double stranded DNA (dsDNA) and thus they can be used as hallmark of the correct performance of the setup by considering $\xi = 50$ nm and $L = 0.34$ nm/bp as reference parameters [Bustamante et al., 1994].

Nonetheless, individual characterization of every molecule in each experiment is highly time-consuming. It is more practical to repeat previous procedure for different molecules to generate a calibration curve of the average force as a function of the position of the magnets and fit the data to an empirical exponential function (Figure 3.8b). This is very convenient since from a given position of the magnets and simple interpolation it is possible to determine the expected applied force during experimental time. However, the error in force (10-15% relative error) reflects the variability in the content of magnetic material of the beads [Lipfert et al., 2009] and the lack of a uniform force along the field of view, specially at low magnifications [De Vlaminck et al., 2012]. Consequently, if precise measurements are required, the force has to be calibrated independently for each molecule.

One of the major advantages of MT compared with other single-molecule techniques is its ability to apply torque on biomolecules. In nick-free DNA molecules, from

now on torsionally-constrained molecules (TC), exerting a torque results in a change of the linking number of the molecule Lk (section 2.3). The introduction of turns changes Lk of the molecule with a relation $\Delta Lk \equiv N$, where N is the number of turns applied with the magnets [Strick et al., 1996]. In the equation 2.1, Lk was described as the sum of twist Tw and writhe Wr. If turns are introduced in the direction of the double helix –counterclockwise turns or by convention positive– the molecule adsorbs the added turns in the form of Tw, while $Wr = 0$. Initally, the applied torque on DNA τ can be semi-quantitatively described from a simple model for twist elasticity as [Strick et al., 2000],

$$\tau(\Delta Tw, F) = \frac{2\pi k_B T C(F)}{L} \Delta Tw \tag{3.32}$$

where L is the contour length of the molecule and C is the torsional modulus, which is a force-dependent parameter [Mosconi et al., 2009]. The torque grows linearly with the number of added turns until it reaches the value

$$\tau_b(F) = \sqrt{2 k_B T \xi F}. \tag{3.33}$$

The linking number such that $\tau(\Delta Tw, F) = \tau_b(F)$ is known as buckling number [Brutzer et al., 2010]. Following addition of turns change the linking number in the form of Wr through the formation of plectonemes, while Tw remains constant. As a consequence, the torque is independent on the number of added turns after the buckling, with a magnitude given by equation 3.33. The formation of plectonemes can be monitored as a linear shortening of the tether extension after the buckling transition (Figure 3.9a). At low forces (< 1 pN) the formation of plectonemes occurs both at positive and negative –clockwise– turns. Conversely, at high forces (> 1 pN) the formation of plectonemes only occurs at positive turns and the introduction of negative turns promotes the melting of the two DNA strands. If the number of negative turns applied is small compared with the linking number of the DNA molecule, the extension remains unaltered with the added turns, resulting in the asymmetric curves shown in Figure 3.9a.

These curves must not be confused with the response observed when beads are tethered to the surface with two or more nucleic acid molecules (Figure 3.9b). In this situation, the first half of turn –either positive or negative– results in a sharp change of extension attributed to the formation of a crossing between the two tethered molecules. Subsequent addition of turns results in a reduction of the extension as a braid where the two DNA molecules are intertwined. This situation persists until the braid buckles forming plectonemes, increasing the change of extension per added turn. The slopes of the first two regimes are dependent on the distance between the two DNA molecules and the length of the tether. Though the DNA:bead ratio can be adjusted to promote the formation of double-tethered beads, the anchoring of DNA molecules to the bead cannot be controlled, resulting in significant differences in the extension-turns profile from bead to bead.

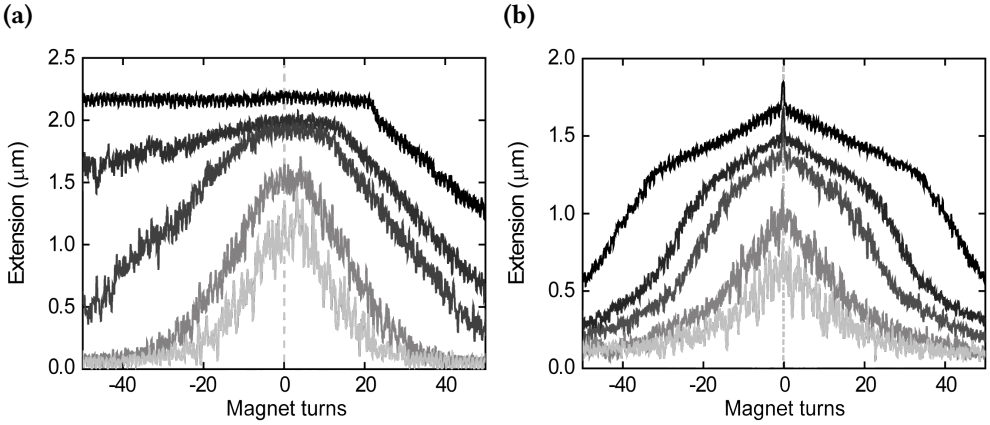

Figure 3.9: Torsional response of tethered DNA molecules. a) Torsionally-constrained DNA. At low forces, adding positive or negative turns results in the formation of plectonemes, monitored from the reduction of the extension after the buckling transition. For forces $>$ 1 pN plectonemes are only formed at positive turns. b) Double-tethered beads symmetric curves. The initial half of turn cross over the two DNA molecules. Aditional turns braid the two DNA molecules until the formation of a plectonemic braid. Stretching forces from light to dark: 0.19, 0.34, 0.85, 1.1 and 2.5 pN.

3.7 Magnetic tweezers improvements

3.7.1 Enhancement of the applied forces

The force exerted on the bead depends on the amount of magnetic material of the bead, and the magnetic field produced by the magnets. The influence of the magnetic field on the beads can be controled during experimental time by modifying the distance between the magnets and the sample. Importanly, following equation 3.1 the force depends on the gradient of the magnetic field. Hence, for a given couple of magnets, the force applied is not determined only by their size and distance to the bead, but also by the shape of the magnetic field generated. This is dependent on the configuration of the magnets in terms of the alignment between them, namely vertical and horizontal (Figure 3.10a). The horizontal configuration is the most common because their positioning on the magnet's head is easier than the vertical configuration. In the vertical configuration, the decay of the force with the distance of the magnets to the surface is lower than in the horizontal configuration and thus if rapid changes of forces are an experimental requirement, the horizontal configuration should be the chosen option. However, for the horizontal configuration the magnet's head have to be fabricated on sweet iron to prevent the action of negative (*pushing*) forces that otherwise will restrict the range of useful magnet positions. The major advantage of the vertical configuration is that if the magnet's head is designed to position the magnets separated by a very small distance ($\approx 200 - 300$ μm) the maximum force that can be applied is significantly higher than that with the same magnets in the horizontal configuration [Lipfert et al., 2009]. Nonetheless,

some concerns must be considered to accurately measure the applied force.

The natural frequency of the DNA-bead system f_0 for a given force is given by

$$f_0(F) = \frac{1}{2\pi}\sqrt{\frac{k}{m}} = \frac{1}{2\pi}\sqrt{\frac{F}{ml}} \tag{3.34}$$

and then caution must be taken when calibrating high forces to prevent camera blurring and aliasing as presented during section 3.5.2.

One option is to increase the acquisition frequency. This is not exempt of drawbacks. Increasing the acquisition frequency –if possible– results in a lower exposure time. The resulting darkened images can lead to tracking errors, especially when measuring at low forces where the focused light used as illumination source is blocked by the magnets. Tracking errors can be reduced by obtaining the LUT at low illumination conditions. If the problem persists there are other alternatives. For instance, increasing the power of the illumination source or activating camera's electronic photomultiplication. Nonetheless, both depend on the capabilities of the MT apparatus and in many cases these will not be feasible.

An apparently straightforward approach is to use long molecules. Equation 3.34 indicates that increasing the length of the molecule reduces the frequency of the system and hence the necessity of working at high sampling frequencies to accurately calibrate the force. However, the fabrication and design of long substrates depends on the availability of the materials, the chemical reactives and the expertise for that purpose, particularly if the tethered substrate of interest is a protein. Moreover, long molecules have several drawbacks. First, long molecules produce huge diffraction rings when extended, reducing the number of beads that can be tracked without interference among them. Second, in MT the apparent *trap* stiffness k depends on the lenght of the substrate $k = F/l$, and then long molecules result in larger fluctuations and therefore in a decrease of the spatial resolution of the system.

A common approach followed has been the use of long molecules for the calibration of forces and interpolating the resulting calibration curve to short substrates [Popa et al., 2016]. This option is not convenient, since the variability in the magnetic moment from bead to bead, as well as the inhomogeneities of the magnetic field result in an error of the applied force estimated from calibration curves.

Fortunately, there are analytical alternatives to determine high forces in short substrates without requiring high-speed tracking. The magnetization of superparamagnetic beads is not isotropic but is biased to the axis of the magnetic field x [Klaue and Seidel, 2009]. Fluctuations of the angle between the anchoring point of the DNA to the bead α (see Figure 3.5) are restricted by the direction of the magnetic field along the x axis. In contrast, fluctuations on the y axis are less constrained by the alignment with the magnetic field and the angle α has freedom to fluctuate. Therefore, in the y axis the effective lenght of tether L can be considered as the sum of the extension of the molecule l and the radius of the bead R, $L = l + R$ [te Velthuis A. J. W et al., 2010; Chen et al., 2011] and applying the same formalism described in

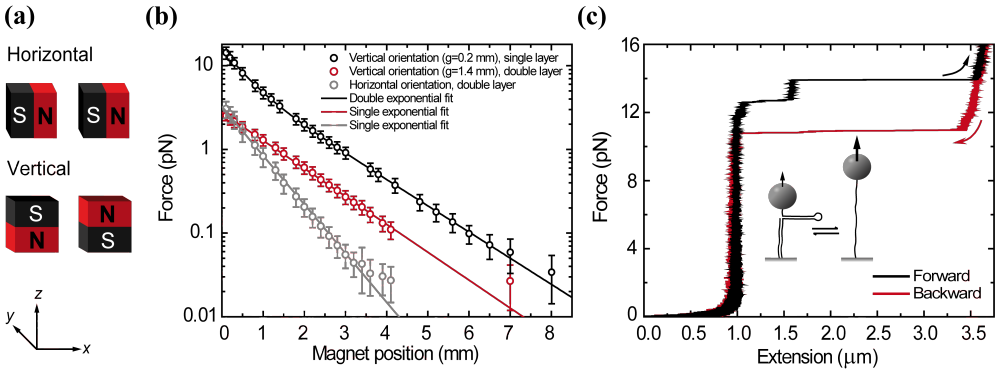

Figure 3.10: Enhancing force range in MT. a) Illustration of magnets' geometries in vertical and horizontal configuration. b) Force calibration in the vertical configuration with gap distances between the magnets of 0.2 mm and single parafilm layer (black), with double parafilm layer and a 1.4 mm gap (red) and for horizontally-aligned magnets (grey). The data was fitted to single or double exponential functions. c) Hairpin unfolding by force used as a proof-of-principle of accurate calibration. For forces > 14 pN the hairpin is unfolded in agreement with the expected unfolding force reported with optical tweezers.

section 3.3 where $k_y = F/(L + R)$, the exerted force F reads

$$F = \frac{k_B T(l + R)}{\langle \delta y^2 \rangle}. \tag{3.35}$$

Recently Daldrop et al. [2015] proposed an elegant calibration method based on the analytical description of the PSD including the angular fluctuations of the bead. Nonetheless, here we attained the analysis to the simpler case of analyzing the force from the mean square displacement on the y axis given by equation 3.35. Following this approximation, it has been calibrated the applied force on 1 μm beads with different magnets configuration and using one or two parafilm layers for the assembly of the liquid chamber (Figure 3.10b). It has been found that the vertical configuration with a gap between magnets of 0.2 mm and using microfluidic cells with a single parafilm layer allows to exert tensions up to $\sim$ 16 pN, compared with a maximum force of $\sim$ 3 pN when the horizontal configuration was used. As a proof of principle, it can be used the unfolding force of a DNA hairpin[2] as a hallmark of bona fide calibration (Figure 3.10c). By stretching the molecule under quasistatic conditions an unfolding force of 14 pN was found, in agreement with the expected unfolding force reported with optical tweezers [Ibarra, 2014] and hence confirming the accuracy of the calibration.

Using single-parafilm layers, 2.8 μm beads and vertically-aligned magnets it is possible to overstretch 2 μm long DNA molecules (applied $F > 65$ pN) and to calibrate the exerted forces employing the methods explained previously. Unfortunately, the weak interaction with the surface between the DIG and the anti-DIG antibody is

[2] A kind gift by Dr. José Morín (Dr. Borja Ibarra's group).

broken at $F \approx 50$ pN and then the acquisition of data is limited at high tension. In the case of requiring to exert forces higher than this, molecules need to be covalently-bound to the surface [Janissen et al., 2014]. Thanks to the advances in the chemical reactions for surface linking and on the methods for accurate force calibration on short substrates here presented, it is expected an expansion of the application of MT for the study of protein folding, where forces in the range between 10 to above 100 pN are required [Rivas-Pardo et al., 2016; Winardhi et al., 2016].

3.7.2 Freely Orbiting Magnetic Tweezers

In a regular MT setup, the configuration of the magnets -either vertically or horizontally-aligned- constrains the angular position of the bead in the xy plane (Figure 3.11a, left). The contribution of the magnetic field on the horizontal plane exerts a torque (equation 3.2) that stringently restricts angular fluctuations to a few degrees around the z axis. In contrast to conventional MT, Jan Lipfert and the group of Nynke Dekker designed a MT system using toroidal magnets, where the magnetic field lines are oriented vertically and then the torque over the xy plane is negligible [Lipfert et al., 2011]. In this MT variant known as freely orbiting magnetic tweezers (FOMT), the bead freely rotates around the z axis driven by thermal fluctuations (Figure 3.11a, right). The rotations of the bead are engaged and restricted by the torsional stiffness of the nucleic acid tether. This coupling between the angular motion of the bead and the DNA allows to extract the torsional properties of the tethered molecules as well as changes in twist induced by drugs or proteins [Li et al., 2015; Lee et al., 2013].

FOMT experiments require a tight alignment of the bead with respect to the magnetic field to suppress x and y contributions to the magnetic moment of the bead. The alignment is possible by displacing the whole magnets' block using the micrometer stage located below the linear motor bracket described in section 3.2.1. Successful alignment results in large amplitude oscillations in the dynamics of the x and y coordinates (Figure 3.11b) which are the product of the circular motion described by the bead (Figure 3.11c). Thereby, it is convenient to describe the angular motion in polar coordinates (r, θ). The average radius of the projected circle on the xy plane, r_0, depends on the applied force and the angle between the anchoring point of the DNA to the bead, α (Figure 3.5). To determine r_0, the xy data is fitted to the equation of a circle

$$r_0 = \pm\sqrt{(x - x_c)^2 + (y - y_c)^2} \tag{3.36}$$

where x_c and y_c are the central coordinates of the circle. The angle α, can be found as $\alpha = \sin^{-1}(r_0/R)$, with R the radius of the bead. Molecules having large off-pole attachment angles α project larger radii and this is desired to facilitate both magnetic field alignment and angular tracking. These molecules can be selected by visual inspection of the amplitude of the rotational motion when magnets turns are applied in conventional MT. Nonetheless, measurement of molecular extension is underrated if the molecule is not perfectly anchored to the south pole of the bead as it was

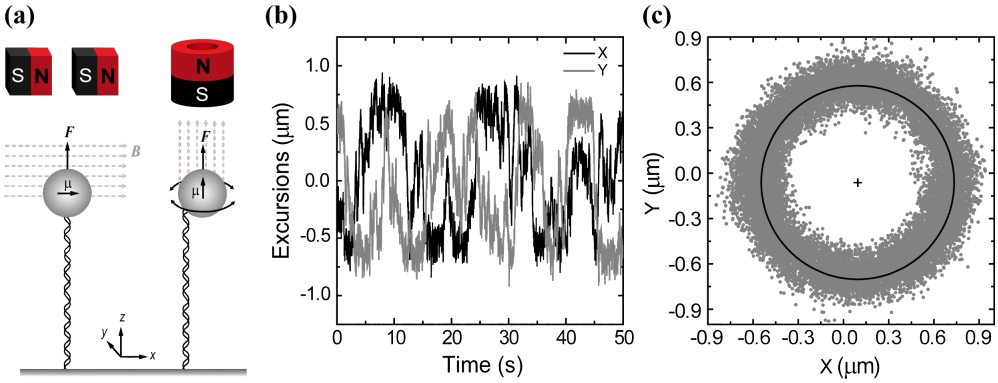

Figure 3.11: Freely orbiting magnetic tweezers. a) In conventional MT (left) the angular position is restricted by the horizontal component of the magnetic moment μ and the magnetic field B (dashed grey lines). Freely orbiting MT (FOMT, right) with z axis alignment of the magnetic field permits free rotation of the bead around the DNA axis. b) Excursions in the x and y coordinates of a bead aligned to the magnetic field in FOMT. c) Circular excursions of the bead and circle fit.

described previously (section 3.4.2). Attending to simple geometrical considerations, the extension of the molecule l including the correction for the anchor point to the DNA can be determined as

$$l = l_m + R(1 - \cos\alpha) \tag{3.37}$$

where l_m is the measured extension. This correction is necessary to accurately calculate the exerted force by FOMT if the radius of the bead is similar to the molecular extension. After corrections, the applied force is determined from the fluctuations of the radius with respect to the equilibrium radius r_0 as

$$F = \frac{k_B T l}{\langle \delta r^2 \rangle} \tag{3.38}$$

where $\langle \delta r^2 \rangle = \langle (r(t) - r_0)^2 \rangle$ is the temporal variance of the radius excursions, obtained from equation 3.36 and the fitting parameters for x_c and y_c.

At a fixed force, the torsional stiffness of the molecule is determined by the angular excursions of the bead. The change of angle $\Delta\theta$ at two consecutive times t can be obtained from the cross-product as

$$\sin^{-1}\Delta\theta(t+1) = \frac{\vec{r}(t) \times \vec{r}(t+1)}{|\vec{r}(t)||\vec{r}(t+1)|} = \frac{x_t y_{t+1} - y_t x_{t+1}}{\sqrt{x_t^2 + y_t^2}\sqrt{x_{t+1}^2 + y_{t+1}^2}}. \tag{3.39}$$

Here, $\vec{r}(t)$ states for center-corrected coordinates, this is, after subtraction of x_c and y_c to the measured x and y coordinates. The absolute angular position can be found as

$$\theta_{t+1} = \theta_{t_0} + \Delta\theta(t) \tag{3.40}$$

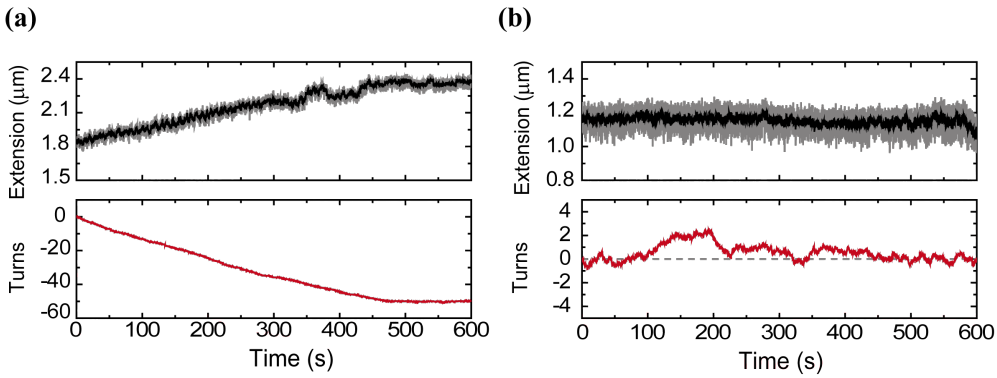

Figure 3.12: Angular excursions in FOMT. a) RecA mediated unwinding monitored by FOMT. The assembly of RecA filaments results in an enlargement of the molecular extension (upper panel) concomitant to DNA unwinding (lower panel). b) Extension and angular fluctuations of a bare torsionally-constrained (TC) DNA molecule. Notice the different scale of the fluctuations. Both traces obtained with the same molecule at $F = 0.8$ pN.

where θ_{t_0} is the initial angular position of the bead determined as

$$\theta_{t_0} = \text{atan2}\left(\frac{y_0}{x_0}\right). \tag{3.41}$$

In this equation, y_0 and x_0 are the coordinates at $t = 0$ and the atan2 function has been rewritten for the range $[0, 2\pi)$. As a proof of principle of the setup and example of application of FOMT to study DNA protein interactions, experiments with *E. coli* RecA –which is commercially available– can be performed. RecA is a protein involved in strand invasion and exchange during homologous recombination [Ohtani et al., 1982]. At moderately acid pH (≈ 6) RecA binds dsDNA and unwinds the double helix. The process of dsDNA unwinding can be monitored with FOMT as a lengthening of the molecule simultaneous to clockwise rotational motion (Figure 3.12a). In contrast, bare TC DNA molecules show a constant extension and a small variability in the angular position of the molecule (Figure 3.12b). In this thesis, FOMT were used to test weather binding and condensation by the DNA-binding protein Spo0J were concomitant to twisting of the molecule, as would be expected for the proposed filamental models described in the literature.

3.7.3 Magnetic Torque Tweezers

In the previous sections it has been shown the simplicity to apply torques on nucleic acids as well as its structural consequences on the tethered molecule. However, measuring the applied torque is not a simple task. Indirect measurements of the applied torque were reported by Mosconi et al. [2009] following the model described by Moroz and Nelson [1997]. In order to obtain precise and direct measurements of the applied torque Lipfert et al. implemented a MT variant capable of working as a torquemeter called Magnetic Torque Tweezers (MTT, Lipfert et al. [2010]). MTT are

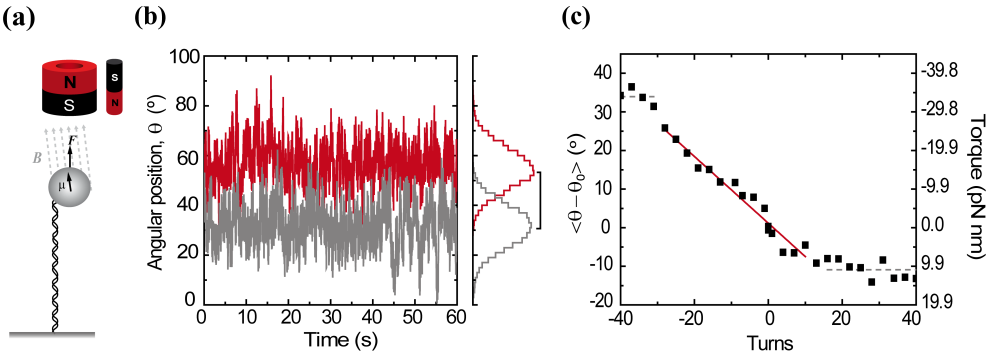

Figure 3.13: Magnetic Torque Tweezers (MTT). a) Scheme of the MTT configuration. Inclusion of a lateral cylindrical magnet to the freely orbiting magnetic tweezers configuration generates a weak angular confinement. b) Angular fluctuations in MTT in absence of turns (gray) and after the introduction of -17 turns (red). The applied force for both situations was 1.5 pN. The histogram shows the distribution of angles and the bar emphasizes the shift in the angular position $\langle \theta - \theta_0 \rangle$. c) Measurement of the torsional stifness of the DNA. From the slope of the torque in the twist regime (red line) a torsional stiffness $C = 78$ nm was determined, in agreement with the reported values at that force (see text).

based on a configuration similar to that shown for FOMT but including a small cylindrical magnet at one side of the main toroidal magnet, which generates a very weak torsional trap on the bead (Figure 3.13a). If the pair of magnets are rotated slowly (0.1 Hz), the nucleic acid accommodates a fraction of the applied turns. This leads to a shift in the average angular position of the twisted molecule (θ) compared with the expected in-plane angle of the relaxed molecule in absence of torque, θ_0 (Figure 3.13b). The applied torque $\tau(\theta)$ is determined by considering a harmonic relation with the shifted angle,

$$\tau(\theta) = k_\theta \langle \theta - \theta_0 \rangle \tag{3.42}$$

where k_θ is the angular trap stiffness, deduced from the energy equipartition theorem as

$$k_\theta = \frac{k_B T}{\langle \delta\theta^2 \rangle}. \tag{3.43}$$

where $\langle \delta\theta^2 \rangle$ is the angular variance, which must be independent of the imposed turns.

MTT can be used to determine the torsional properties of nucleic acids, both in absence or presence of proteins. By equaling equation 3.42 with equation 3.32, the torsional stiffness of the molecule C can be determined from the slope of the torque-turns plot before the buckling transition (Figure 3.13c). The torsional stiffness of the DNA is a force-dependent parameter reaching a saturation value of ≈ 95 nm for forces > 4 pN [Mosconi et al., 2009]. MTT have been used to precisely determine this force-dependent relation of the torsional stiffness as well as the influence of the ionic conditions [Kriegel et al., 2017]. Interestingly, while the torsional stiffness of the DNA is largely independent on the ionic conditions of the environment,

this parameter can be strongly influenced by protein binding. Thereby, MTT can be used to characterize structural properties of protein-DNA complexes. Particularly, measurements of torque and torsional stiffness have been used as a probe of the mechanochemical features and the degree of coverage of proteins forming filaments on nucleic acids, as is the case of DNA repair proteins [Lee et al., 2013; Lipfert et al., 2010].

3.8 Concluding remarks and future directions

In this chapter, it has been presented a detailed description of a MT setup, including both hardware and software methods. Examples of applications have been briefly introduced with didactic purposes. A clear frame of its application to the study of DNA protein interactions will be presented in the following chapters of this book. In addition, guidelines for straightforward and powerful improvement of the capabilities of the system have been discussed. The technique is under constant development and it benefits from the advances in new technologies. For instance, the increasing number of cores in graphic cards and the fast acquisition frequency of recent CMOS (complementary metaloxidesemiconductor) cameras are allowing to extract highly parallel measurements even at high frequencies. This enables sampling the dynamics with high resolution and collecting impressive statistics [Dulin et al., 2015a,b]. More sophisticated improvements include the combination of MT with other single-molecule techniques and in the last years one of the main trend has been the combination of MT with fluorescence microscopy [Swoboda et al., 2014]. This allows to correlate protein binding and its mechanical implications independently of protein activity, or to monitor the effect of tension on the structure of nucleic acids [Schwarz et al., 2013; van Loenhout et al., 2012b]. The combination of these two highly-parallelizable techniques expands the experimentally accessible information and allows the design of complex experiments to understand the underlying complicated mechanisms of biological systems.

References

Beausang, J. F., Zurla, C., Finzi, L., Sullivan, L., and Nelson, P. C. Elementary simulation of tethered Brownian motion. *Am. J. Phys.*, 75(6):520–3, 2007.

Berghuis, B. A., Dulin, D., Xu, Z. Q., van Laar, T., Cross, B., Janissen, R., Jergic, S., Dixon, N. E., Depken, M., and Dekker, N. H. Strand separation establishes a sustained lock at the Tus-Ter replication fork barrier. *Nat. Chem. Biol.*, 11 (8):579–85, 2011.

Brutzer, H., Luzzietti, N., Klaue, D., and Seidel, R. Energetics at the DNA supercoiling transition. *Biophys. J.*, 98(7): 1267Ű1276, 2010.

Bustamante, C., Marko, J. F., Siggia, E. D., and Smith, S. Entropic elasticity of lambda-phage DNA. *Science*, 265(5175): 1599–1600, 1994.

Chatfield, C. *The analysis of time series: an introduction.* CRC Press, 6th edition, 2013.

Chen, H., Fu, H., Zhu, X., Cong, P. Nakamura, F., and Yan, J. Improved high-force magnetic tweezers for stretching and refolding of proteins and short DNA. *Biophys. J.*, 100(2):517–23, 2011.

Czerwinski, F., Richardson, A. C., and Oddershede, L. B. Quantifying noise in optical tweezers by Allan variance. *Opt. Express*, 17(15):13255–69, 2009.

Daldrop, P., Brutzer, H., Huhle, A., Kauert, D. J., and Seidel, R. Extending the range for force calibration in magnetic tweezers. *Biophys. J.*, 108(10): 2550–61, 2015.

De Vlaminck, I., Henighan, T., van Loenhout, M. T., Pfeiffer, I., Huijts, J., Kerssemakers, J. W., Katan, A. J., van Langen-Suurling, A. van der Drift, E., Wyman, C., and Dekker, C. Highly parallel magnetic tweezers by targeted DNA tethering. *Nano Lett.*, 11(12):5489–93, 2011.

De Vlaminck, I., Henighan, T., van Loenhout, M. T., Burnham, D. R., and Dekker, C. Magnetic forces and DNA mechanics in multiplexed magnetic tweezers. *PLoS One*, 7(8):e41432, 2012.

Dulin, D., Cui, T. J., Cnossen, J., Docter, M. W., Lipfert, J., and Dekker, N. H. High spatiotemporal-resolution magnetic tweezers: Calibration and applications for DNA dynamics. *Biophys. J.*, 109(10):2113–25, 2015a.

Dulin, D., Vilfan, I. D., Berghuis, B. A., Hage, S., Bamford, D. H., Poranen, M. M., Depken, M., and Dekker, N. H. Elongation-competent pauses govern the fidelity of a viral RNA-dependent RNA polymerase. *Cell Rep.*, 10(6): 983Ű992, 2015b.

Gelles, J., Schanapp, B. J., and Sheetz, M. P. Tracking kinesin-driven movements with nanometre-scale precision. *Nature*, 331(66615):450–453, 1988.

Gosse, C. and Croquette, V. Magnetic tweezers: micromanipulation and force measurement at the molecular level. *Biophys J.*, 82(6):3314–29, 2002.

Hell, S., Reiner, G., C. Cremer, C., and Stelzer, E. H. K. Aberrations in confocal fluorescence microscopy induced by mismatches in refractive index. *J. Microsc.*, 169(3):391–405, 1993.

Ibarra, B. Personal communication, 2014.

Janissen, R., Berghuis, B. A., Dulin, D., Wink, M., van Laar, T., and Dekker, N. H. Invincible DNA tethers: covalent DNA anchoring for enhanced temporal and force stability in magnetic tweezers experiments. *Biophys. J.*, 42(18):e137, 2014.

Klaue, D. and Seidel, R. Torsional stiffness of single superparamagnetic microspheres in an external magnetic field. *Phys. Rev. Lett.*, 102(2):028302, 2009.

Koster, D. A., Croquette, V., Dekker, C., and Shuman, S. Dekker, N. H. Friction and torque govern the relaxation of DNA supercoils by eukaryotic topoisomerase IB. *Nature*, 434(7033):671–674, 2005.

Kriegel, F., Ermann, N., Forbes, R., Dulin, D., Dekker, N. H., and Lipfert, J.

Probing the salt dependence of the torsional stiffness of DNA by multiplexed magnetic torque tweezers. *Nucleic Acids Res.*, 45(10):5920–5929, 2017.

Lansdorp, B. M. and Saleh, O. A. Power spectrum and allan variance methods for calibrating single-molecule video-tracking instruments. *Rev. Sci. Instrum.*, 83(2):025115–10, 2012.

Lee, M., Lipfert, J. Sanchez, H., Wyman, C., and Dekker, N. H. Structural and torsional properties of the RAD51-dsDNA nucleoprotein filament. *Nucleic Acids Res.*, 41(14)::7023–30, 2013.

Li, W., Wong, W. J., Lim, C. J., Ju, H. P., Li, M., Yan, J., and Wang, P. Y. Complex kinetics of DNA condensation revealed through DNA twist tracing. *Phys. Rev. E. Sta.t Nonlin. Soft Matter Phys.*, 92(2): 022707, 2015.

Lipfert, J., Haom, X., and Dekker, N. H. Quantitative modeling and optimization of magnetic tweezers. *Biophys. J.*, 96(12):5040–5049, 2009.

Lipfert, J., Kerssemakers, J. W., Jager, T., and Dekker, N. H. Magnetic torque tweezers: measuring torsional stiffness in DNA and RecA-DNA filaments. *Nat. Methods*, 7(12):977–80, 2010.

Lipfert, J., Wiggin, M., Kerssemakers, J. W., Pedaci, F., and H., D. N. Freely orbiting magnetic tweezers to directly monitor changes in the twist of nucleic acids. *Nat. Commun.*, 2:439, 2011.

Moroz, J. D. and Nelson, P. Torsional directed walks, entropic elasticity and DNA twist stiffness. *Proc. Nat. Acad. Sci. USA.*, 94(26):14418Ű22, 1997.

Mosconi, F., Allemand, J. F., Bensimon, D., and Croquette, V. Measurement of the torque on a single stretched and twisted DNA using magnetic tweezers. *Phys. Rev. Lett.*, 102(7):078301, 2009.

Ohtani, T., Shibata, T., Iwabuchi, M., Watabe, H., Iino, T., and Ando, T. Atp-dependent unwinding of double helix in closed circular DNA by RecA protein of E. coli. *Nature*, 299(5878):86–9, 1982.

Popa, I., Rivas-Pardo, J. A., Eckels, E. C., Echelman, D. J., Badilla, C. L., Valle-Orero, J., and Fernández, J. M. A HaloTag anchored ruler for week-long studies of protein dynamics. *J. Am. Chem. Soc.*, 138(33):10546–53, 2016.

Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P. *Numerical Recipes in C. The art of scientific computing.* Cambridge University Press, 2nd edition, 2009.

Rivas-Pardo, J. A., Eckels, E. C., Popa, I., Kosuri, P., Linke, W. A., and Fernández, J. M. Work done by titin protein folding assists muscle contraction. *Cell Rep.*, 14 (6):1339–47, 2016.

Schwarz, F. W., Tóth, J., van Aelst, K., Cui, G., Clausing, S., Szczelkun, M. D., and Seidel, R. The helicase-like domains of type III restriction enzymes trigger long-range diffusion along DNA. *Science*, 340(6130):353–6, 2013.

Seidel, R., van Noort, J., van der Scheer, C., Bloom, J. G., Dekker, N. H., Dutta, C. F., Blundell, A., Robinson, T., Firman, K., and Dekker, C. Real-time observation of DNA translocation by the type I restriction modification enzyme EcoR124I . *Nat. Struct. Mol Biol.*, 11(9): 838–43, 2004.

Smith, S. B., Finzi, L., and Bustamante, C. Direct mechanical measurements of the elasticity of single DNA molecules by using magnetic beads. *Science*, 258 (5085):1122–6, 1992.

Strick, T., Allemand, J., Croquette, V., and Bensimon, D. Twisting and stretching single DNA molecules. *Prog. Biophys. Mol. Biol*, 74(1-2):115–40, 2000.

Strick, T. R., Allemand, J. F., Bensimon, D., Bensimon, A., and Croquette, V. The elasticity of a single supercoiled DNA molecule. *Science*, 271(5257):1835–7, 1996.

Swoboda, M., Grieb, M. S., Hahn, S., and Schlierf, M. *Measuring Two at the Same Time: Combining Magnetic Tweezers with Single-Molecule FRET*, pages 253–276. Springer Basel, Basel, 2014. ISBN 78-3-0348-0856-9.

te Velthuis A. J. W, Kerssemakers, J. W. J., Lipfert, J., and H., D. N. Quantitative guidelines for force calibration through spectral analysis of magnetic tweezers data. *Biophys J.*, 99(4):1292–302, 2010.

Tipler, P. A. and Mosca, G. P. *Física para la ciéncia y la tecnología*. Reverté, 6a edition, 2010.

van Loenhout, M. T., Kerssemakers, J. W., De Vlaminck, I., and Dekker, C. Non-bias-limited tracking of spherical particles, enabling nanometer resolution at low magnification. *Biophys J.*, 102(10):2362–71, 2012a.

van Loenhout, M. T. J., de Grunt, M. V., and Dekker, C. Dynamics of DNA supercoils. *Science*, 338(6103):94–97, 2012b.

van Oene, M. M., Dickinson, L. E., Pedaci, F., Köber, M., Dulin, D., Lipfert, J., and Dekker, N. H. Biological magnetometry: torque on superparamagnetic beads in magnetic fields. *Phys. Rev. Lett.*, 114(21):218301, 2015.

Vilfan, I. D., Lipfert, J., Koster, D. A., Lemay, S. G., and Dekker, N. H. Magnetic tweezers for single-molecule experiments. In *Handbook of Single-Molecule Biophysics*, pages 371–395. Springer US, 2009.

Winardhi, R. S., Tang, Q., Chen, J., Yao, M., and Yan, J. Probing small molecule binding to unfolded polyprotein based on its elasticity and refolding. *Biophys. J.*, 111(11):2349–2357, 2016.

Wong, W. P. and Halvorsen, K. The effect of integration time on fluctuation measurements: calibrating an optical trap in the presence of motion blur. *Opt. Express.*, 14(25):12517–31, 2006.

Part II

The relation between bacterial chromosome condensation and segregation: the protein ParB as a model

4

DNA condensation by the bacterial protein ParB

Chapter overview. The condensation and dynamic re-organization of the chromosome is crucial to the cell cycle of all living organisms. This process is tightly linked to chromosome segregation and it is essential for the accurate transmission of the genetic information to the offspring. In Bacteria, the DNA binding protein ParB was originally described as a component of the segregation apparatus, forming a centromere-like structure around its specific DNA binding sequence *parS*. This chapter explores an interesting novel property of ParB in condensing DNA molecules. The condensed DNA structures are not well ordered and can be inferred to be formed by many looping interactions between neighboring DNA segments. Consistent with this view, ParB is also able to stabilise writhe in single supercoiled DNA molecules and to bridge segments from two different DNA molecules. Our magnetic tweezers experiments provide no evidence for the promotion of non-specific DNA binding and/or condensation events by the presence of *parS* sequences. The implications of these findings for the condensation and organization of the bacterial chromosome are discussed.

This chapter has been published by James A. Taylor, Cesar L. Pastrana,..., Fernando Moreno-Herrero, and Mark S. Dillingham in *Nucleic Acids Research*, 2015 43(2): 719-31.

4.1 Introduction

4.1.1 DNA condensation

In all domains of life, the genome is compacted into a high-order structure with a high density of proteins bound. The processes that allow the compaction of the DNA into a reduced volume are known as condensation. The DNA is a highly charged polymer and consequently it is relatively stiff. Simple calculations allow determining the expected volume occupied by certain genome. Considering the worm-like chain model explained in the section 2.4.2, it is possible to quantify an average size of the DNA by means of the radius of gyration, R_g. This variable measures the mean distance to the center of mass of the polymer, and can be extracted from the

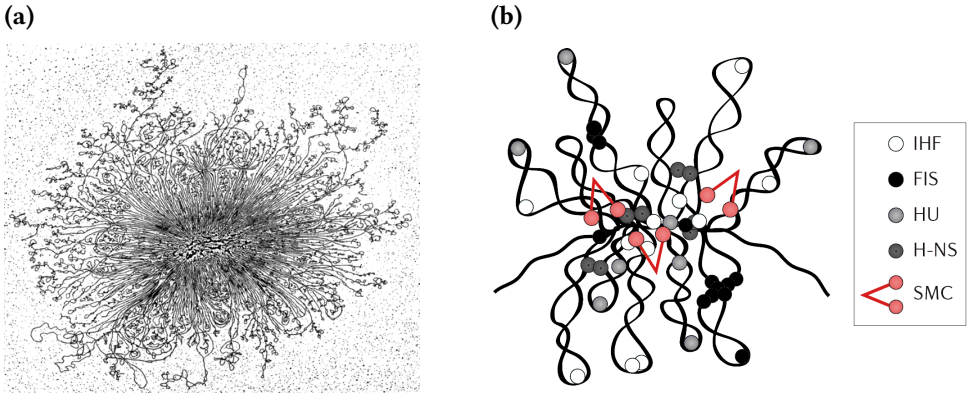

Figure 4.1: Condensation and organization of the bacterial chromosome. a) Electron microscopy picture of an *E. coli* bacterium surrounded by its genome. Hypherosmotic treatment leads to the ejection of the DNA from the cell. Image by Dr. Ruth Kavenoff from Griswold [2008]. b) Schematic representation of nucleoid-associated proteins in collaboration with SMC complexes to supercoil and condense the bacterial chromosome. Adapted from Wang et al. [2013].

expresion $\sqrt{\langle R_g^2 \rangle} = \sqrt{L\xi/3}$, with L and ξ as the contour and persistence length respectively [Phillips et al., 2009]. The genome of *Bacillus subtilis* has 4.2 Mbp ($L = 1.5$ mm), resulting in $R_g \approx 5$ μm. Assuming that the chromosome adopts a sphere-like structure, this results in a volume of 525 μm^3. The cell volume for that particular bacteria has been estimated in 4.6 μm^3 [Yu et al., 2013], and consequently the volume occupied by DNA must be reduced one hundred times to fit inside the cell. The relation between the size of the genome and the cell is stressed in Figure 4.1a. This huge degree of compaction is achieved through DNA supercoiling (see section 2.3) and by the action of specialized DNA-binding proteins. In Bacteria, the condensation of the genome is dependent on the so-called nucleoid-associated proteins [Wang et al., 2013]. These proteins condense the chromosome by forming loops and bending the DNA [Dame, 2005]. Chromosome condensation is a highly regulated dynamical process which involves additional proteins (Figure 4.1b). Among them, a key protein complex is the Structural Maintenance of Chromosome (SMC). SMCs are part of the condensins family, large structurally-conserved protein complexes involved in DNA condensation and chromosome segregation which are also present in eukaryotes [Hirano, 2009]. In Bacteria, SMC proteins dimerize forming a ring-like structure in complex with the accesory proteins ScpA and ScpB [Hirano, 2006]. It has been proposed that SMC encloses DNA molecules, entrapping topological domains between its two arms in a ATP-regulated manner [Britton et al., 1998; Thanbichler et al., 2005]. However, the precise mode of interaction of the SMC complexes with the DNA as well as the mechanisms leading to condensation and domain entrapment remains unknown.

4.1.2 Chromosome segregation

The correct partitioning of the chromosomes is essential for the cell, both as individual and as species. Chromosome segregation requires to be highly organized and coordinated with DNA replication to prevent the loss of genetic information from generation to generation. Although the process of chromosome segregation is relatively well-known in eukaryotes, the mechanisms and the machinery involved in chromosomal segregation remains unclear for prokaryotes. The hypothesis of a mitosis-like apparatus is the most consistent, with the origin of replication (*ori*) and proximal regions having a similar role to that of the centromere in eukaryotic cells [Wang et al., 2013]. The active segregation of the chromosome in Bacteria is dependent on a partition system consisting on a centromere-like DNA sequence, a centromere-binding protein and a *motor* protein, typically an ATPase [Schumacher, 2008].

Plasmids are extrachromosomal DNA molecules that are replicated independently of the main chromosome and confere new functions to the host. Nonetheless, plasmids have been used as a model of bacterial chromosome segregation. Plasmid partition systems can be divided into different evolutionary unrelated families classified attending to the structure of the NTPase [Gerdes et al., 2000], namely Type I (encompassing Types Ia and Ib), Type II and Type III [Oliva, 2016]. In all the variants, plasmid segregation is dependent on the binding of a set of centromere-binding proteins to the centromere sequence. Then, the nucleoprotein complex recruits the motor protein, and this exerts the driving force to segregate the chromosome [Schumacher, 2011]. There are different working models for chromosome segregation attending to the type of the partitioning system involved [Gerdes et al., 2000].

In all the currently proposed models, plasmid segregation is dependent on the polymerization of the motor protein forming a filament-like structure. In the Type III system, a dynamic filament transports the bacterial genome following a similar mechanism to that of tubulin during mitosis in eukaryotes. The model proposed for the Type II, envisions the generation of a filament joining the two newly synthesized plasmids. The filament grows by insertion of monomers in between the filament, and thus both genomes are pushed apart upon polymerization. In the Type I partition system the motor proteins form a filament, but polymerization occurs on the plasmid. The polymerization proceeds until interaction with the centromere-binding protein, where the stimulation of NTPase activity results in the depolymerization of the filament. Nonetheless, is not known how this process results in chromosome segregation. Importantly, the Type I system has been considered a model system for bacterial chromosome segregation, to which the ParABS system belongs.

4.1.3 The ParABS system

The Type I partition system ParABS was discovered and originally defined as the DNA segregation apparatus for low copy number plasmids [Gerdes et al., 2010]. Today, Type I is known as the most widely present system for bacterial chromsome

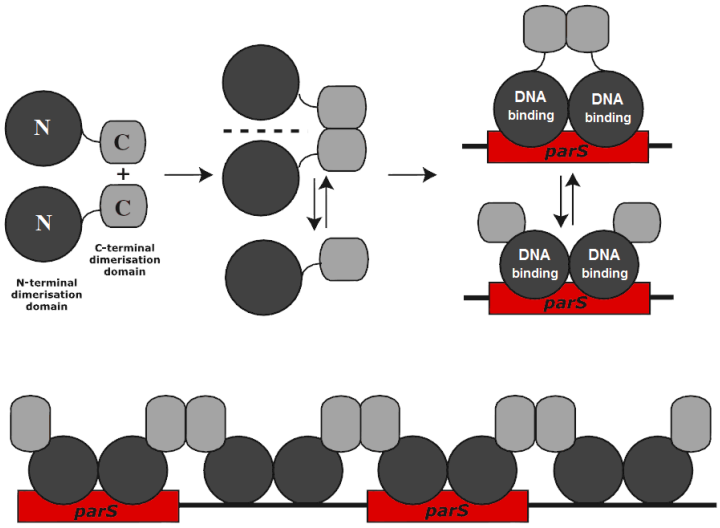

Figure 4.2: ParB spreading model from *parS*. ParB dimerizes at the C-terminal domain and binds DNA as dimer recognizing the *parS* sequence by the helix-turn-helix motif located at the N-terminal domain. DNA binding promotes dimerization trough the N-terminal domain and the recruitment of additional ParB units by dimerization with the C-terminal domain. Adapted from Leonard et al. [2004].

segregation [Wang et al., 2013]. The ParABS system comprises ParB, a DNA binding protein, its binding sequence termed *parS*, and ParA protein. ParA is a DNA binding protein with ATPase activity. The ParB-*parS* complex forms a centromere-like structure which is then recognized and segregated by ParA. A homologous apparatus is important for chromosome segregation during cell division or sporulation in many bacteria including *Bacillus subtilis* [Mierzejewska and Jagura-Burdzy, 2012; Yamaichi and Niki, 2000]. In *B. subtilis* there are eight functional *parS* sequences (consensus sequence TGTTTCACGTGAAACA) located near the *ori*, resulting in an acumulation of ParB –also called Spo0J in *B.subtilis*– at these regions [Lewis and Errington, 1997; Lin and Grossman, 1998]. ParB binds non-specifically to DNA sequences in the vicinity of *parS* sites for several kilobases [Lin and Grossman, 1998; Breier and Grossman, 2007]. Based on structural data, Leonard et al. [2004] proposed a model where after dimerization through the C-terminal domain ParB binds specifically to the *parS* sequence (Figure 4.2). The specific binding is dependent on the helix-turn-helix motif located at the N-terminal domain. The N-terminal domain forms a secondary dimerization site necessary for DNA binding. After binding the *parS* sequence, the dimerization domain at the C-terminal interacts with the C-terminal domain of a different ParB protomer, promoting the binding to adjacent non-specific DNA sites. Therefore, the *parS* sequence acts as a nucleation site for the formation of a ParB filament that spreads to non-specific DNA regions. This phenomenon has been shown to be important for the biological function in several cases [Breier and Grossman, 2007; Rodionov and Yarmolinsky, 2004; Murray et al., 2006]. It is known that null mutants in ParB gene are lethal or produce anucleate cells depending on the species

[Mohl et al., 2001; Bartosik et al., 2009]. In *B. subtilis* the spreading mutation R80A results in an abnormal nucleoid and are defective in sporulation [Autret et al., 2001]. In fact, non-specific binding from *parS* sequences is a general feature of ParB proteins [Schumacher, 2008]. However, the mechanism of ParB spreading from *parS* sequences and the implications of spreading to hallmark chromosome segregation are not understood.

Remarkably, the interaction between *B. subtilis* ParB and *parS* is implicated in the recruitment of the SMC complex to the chromosome, providing an intriguing structural link between chromosome partitioning and condensation [Gruber and Errington, 2009; Sullivan et al., 2009]. The details of how ParB proteins assemble to form higher order *segrosome* complexes with their DNA targets remain to be determined and may differ, even between ParB proteins that are related at the level of primary structure [Schumacher, 2008]. In addition, how ParB triggers the loading of SMC proteins and how the activity of the condensin is enhanced by ParB proteins around the *ori* remains unknown.

Electrophoretic mobility shift assays (EMSAs) on short oligonucleotide substrates, as well as footprinting assays, have demonstrated specific binding of *parS* by *B. subtilis* ParB *in vitro* but do not provide evidence for the lateral spreading that had been expected based on the *in vivo* behaviour [Breier and Grossman, 2007; Murray et al., 2006]. Moreover, several previous studies using EMSAs with longer DNA substrates showed an apparent *laddering* effect that was consistent with non-specific coating of DNA by ParB, but did not directly address the role of *parS* sequences in (potentially) nucleating these structures [Murray et al., 2006; Sullivan et al., 2009]. Recently, single molecule imaging studies have revealed that several different ParB proteins are able to condense flow-stretched DNA, and it has been suggested that this is the result of protein-mediated bridges between distant non-specific DNA segments [Graham et al., 2014]. The possible role of *parS* in this activity was unclear because the study focussed on substrates lacking specific sites, and ParB localization at *parS* was not observed. Based in part upon this work, modelling of ParB binding to DNA has suggested a *two-state nucleation model* for ParB binding and condensation of DNA around *parS* [Broedersz et al., 2014]. In this model, a combination of 3D bridging and lateral spreading protein-protein interactions form a network of ParB protomers that together condense DNA, and it is proposed that such networks are most efficiently nucleated by *parS*-bound ParB molecules.

In this chapter, we show using EMSA assays that *Bacillus subtilis* ParB binds to DNA both specifically and non-specifically. Magnetic tweezers (MT) show that ParB is capable of packaging non-specific DNA with a low condensing force of $\approx$2 pN. This condensation activity is reversible and due to the formation of loops between non-specific DNA segments that are bridged *in cis* by multiple ParB proteins. Bridging *in trans* was also observed in experiments involving multiple-tethered DNA molecules. The large nucleoprotein complexes are presumably formed by a combination of ParB-ParB interactions together with non-specific ParB-DNA interactions. We speculate that non-specific binding may occur at an unidentified site on the protein, such that the nucleoprotein condensates could be anchored around *parS* by an

independent and highly specific binding locus associated with the helix-turn-helix motifs.

4.2 Materials and Methods

4.2.1 Protein purification

Untagged ParB was purified by James A. Taylor, PhD from the group of Prof. Mark S. Dillingham (University of Bristol). The detailed procedure can be found in Taylor et al. [2015].

4.2.2 Electrophoretic mobility-shift assays (EMSA)

ParB was incubated[1] with 20 nM 147 bp DNA substrate (consisting of 19 nM cold DNA spiked with 1 nM radiolabelled DNA), 50 mM Hepes-KOH, 100 mM KCl, 2.5 mM $MgCl_2$, 0.1 mg/mL BSA, 1 mM DTT and 2.5 % (v/v) Ficoll in a 25 µl reaction volume. Samples were incubated at room temperature for 30 minutes followed by 5 minutes on ice. Complexes were resolved on 6% acrylamide/bis-acrylamide (29:1) gels in 90 mM Tris, 150 mM Boric acid (final pH 7.5). Gels were supplemented with 2.5 mM $Mg(Ac)_2$ (TBM gel), and were pre-run at 150 V, 4°C for 30 minutes. Ten microliters of each sample were loaded and the gels run at 150 V, 4°C for 1 hour. The gels were then dried under vacuum and visualised by exposure to a phosphor screen, which was subsequently scanned by a phosphorimager (Typhoon FLA 9500, GE Healthcare). Gels were quantified using ImageQuant (GE Healthcare) and each band expressed as a percentage of the sum of all bands. The non-specific binding data was sigmoidal, suggesting positive co-operativity, and was treated semi-quantitatively by fitting to a Hill equation.

$$p = \frac{[\mathrm{ParB_2}]^h}{K_A^h + [\mathrm{ParB_2}]^h}, \tag{4.1}$$

where p is the measured binding signal, h is the Hill coefficient and K_A is the concentration of protein producing half occupation. The errors reported are the standard deviation of three independent data sets.

4.2.3 DNA substrates

Substrates for magnetic tweezers were prepared by PCR amplification of the pET28a plasmid containing the *parS* sequence using a high fidelity DNA polymerase (Phusion Polymerase, New England Biolabs) with NotI and XhoI-containing primers. The NotI and XhoI sites were removed from the pET28a plasmid by site directed mutagenesis[1]. The non-specific substrate is a mutated version of the *parS* sequence were the GC content was kept constant but that prevented for ParB specific binding (CGTGCCCAGGGAGACA). The PCR product (6 kbp) was cut with NotI and

[1]The biochemical aspects of this work regarding protein purification, cloning, site directed mutagenesis and EMSA assays performed by James A. Taylor, PhD.

XhoI (both from NEB). Handles (1 kbp) were prepared by PCR amplification from pSP73-JY0 [Fili et al., 2010] including biotine-16-dUTP or digoxigenin-11-dUTP in the reaction (both from Roche Diagnostics). This was followed by restriction with NotI (biotin handle) or XhoI (digoxigenin handle). Finally, handles and the central fragment were ligated by incubation with T4 DNA ligase (New England Biolabs) overnight at 16°C followed by heat inactivation for 20 min at 65°C. DNAs were never exposed to intercalant dyes or UV radiation during their production and were stored at 4°C. The molecules were ready to use in MT. This procedure allows the obtention of torsionally-constrained molecules with a $\approx$ 30 % yield.

DNA molecules for competitition assays (50 bp) were fabricated by annealing of complementary oligos in a 1:1 ratio. The annealing was achieved by decreasing the temperature from 98°C to 25°C at a rate of 1°C/s in a buffer composed of 10 mM Tris-HCl pH 7.5, 100 mM NaCl and 1 mM ethylenediaminetetraacetic acid (EDTA).

4.2.4 Magnetic tweezeers conditions

We used the home-built setup similar to that described in section 3.2.1 but operating with a CCD camera with a maximum acquisition frequency of 120 Hz (TM-6710CL, Pulnix Inc USA). Data were acquired at 60 Hz.
The experiments were carried out in a reaction buffer composed of 100 mM NaCl, 50 mM Hepes-KOH pH 7.5, 4 mM $MgCl_2$, 1 mM DTT, 100 µg/ml BSA and 0.1% Tween-20. Molecules of the proper length were first selected by their force-extension response. Double or torsionally constrained molecules were selected acording to their characteristic magnet rotation-extension relation (section 3.6) and discarded unless indicated otherwise. The protein was flown into the liquid chamber and incubated for a couple of minutes before starting the measurements. Incubation of 5 M NaCl followed by rinse with several cell volumes of ParB reaction buffer induced protein release from the DNA. This allowed us to measure condensation and decondensation events using the same DNA molecules. Some experiments were conducted by flowing proteins or DNAs during measurement time. In these cases the flow was set to 18 µl/min to prevent tracking errors. A horizontal bar in the figure indicates the starting and ending times of the flow.

Calculation of the duration and extent of condensation
The final extension l_f was the mean extension in a time window of two seconds at the end of a condensation event. We considered that a condensation event finished when there are no visible changes in extension for several tens of seconds. The condensation time t_c was determined as the time to reach an extension l given by $l = l_f - 2\sigma_l$, where σ_l is the standard deviation of the extension at the end of the condensation process. This criterion eliminated small condensation events occurring after long periods of time that would otherwise skew the time distribution to longer values.

Force-extension curves
For each DNA molecule, a value of the extension at a given force was calculated. Then, a mean value and standard deviation were determined for multiple DNA molecules. In the absence of protein, data were fitted to the worm-like chain model [Bouchiat et al., 1999]. From the fits, we obtained the contour length (L) of the molecule, which was subsequently used to normalize the data. We measured a mean L of 2.2 μm. The molecule previously characterized in the absence of protein was then studied in the presence of 1 μM $ParB_2$. ParB force extension data was calculated as above but was not fitted to any model.

Condensation force
To determine the force at which the process of condensation starts, the magnet was alternated between positions that apply a large stretching force (3.9 pN) or a variable smaller force that was gradually decreased. The low force was maintained for 35 s and the mean extension for the first (E_0) and the last five seconds (E_f) of the measurement determined. The condensation force is defined as the critical force that satisfies the criterion, $E_f < E_0 - 10\sigma_0$, where σ_0 is the standard deviation of E_0 at 1 Hz bandwidth. This threshold value ($10\sigma_0$) allowed us to discriminate between real condensation events and oscillations that are sometimes visible at forces near condensation.

Freely Orbiting Magnetic Tweezers
Freely orbiting magnetic tweezers (FOMT) were performed at a stretching force of 0.45 pN on *parS*-containing substrates. Data recording was initiated immediately after protein injection. The details of the FOMT technique can be found in the Magnetic Tweezers chapter (section 3.7.2).

4.3 Results

4.3.1 Non-specific DNA binding by ParB leads to condensation

It has been shown previously that *B. subtilis* ParB is associated with non-specific DNA for several kilobases either side of its specific DNA binding sites *in vivo* [Breier and Grossman, 2007; Murray et al., 2006]. This has led to the hypothesis that the *parS* sequence acts as a nucleating site from which ParB spreads either laterally [Murray et al., 2006] or by the formation of loops [Graham et al., 2014]. However, *in vitro* studies of *B. subtilis* ParB binding and spreading have produced contradictory and confusing results. For example, studies which used EMSAs to investigate binding to long DNAs produced a ladder of ParB:DNA complexes, apparently consistent with coating of non-specific DNA, whilst DNAse I footprinting showed only protection of *parS* with no evidence for spreading to non-specific sites [Sullivan et al., 2009; Murray et al., 2006]. Furthermore recent studies on ParB binding to flow-stretched lambda DNA were unable to visualise binding of fluorescently labelled ParB to *parS*, but could observe rapid non-specific binding which resulted in DNA condensation [Graham et al., 2014].

To investigate binding of ParB to specific and non-specific DNA, we first devel-

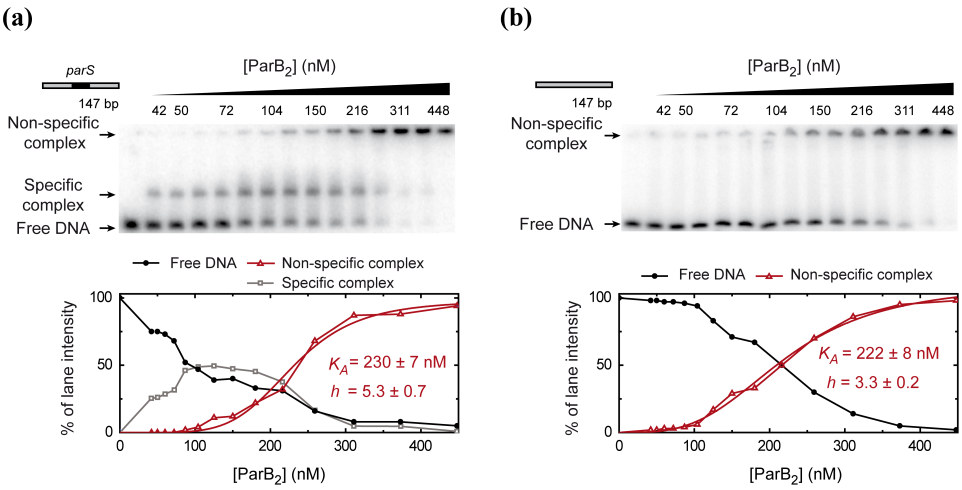

Figure 4.3: Specific binding of ParB to the *parS* sequence. Electrophoretic mobility shift assay of ParB binding to a radiolabelled 147 bp substrate in a magnesium acetate containing gel-running buffer. a) Titration of ParB on DNA containing a single *parS* site in the centre. b) ParB titration on an equivalent substrate that is lacking a *parS* site. The species assigned as specific and non-specific complexes are labelled. The lower panels show the quantification of the gels revealing a highly pattern for non-specific binding. These data were fit to the Hill equation to yield the values shown. EMSA assays performed by James Taylor, PhD (Prof. Mark Dillingham, University of Bristol)

oped a protocol for the purification of native (untagged) ParB protein, and then revisited the use of EMSA assays using a variety of different buffer compositions for resolving nucleoprotein complexes in the gels. We subsequently found that, using a gel running buffer with magnesium acetate instead of EDTA, we were able to clearly reveal the specific binding of ParB to *parS* on long DNA molecules (Figure 4.3). Under these conditions, the *parS* substrate shows a shift at low [ParB] to form a single well-defined complex that runs into the gel, and a supershift of this material to a position close to the wells of the gel at elevated [ParB]. When the control DNA is used, the well-resolved intermediate band is not observed at all, whereas the shift to the wells still occurs, and does so at concentrations of ParB similar to those for the *parS*-containing substrate. We interpret the intermediate band as a specific complex of ParB bound to the *parS* sequence and the dissociation constant for this interaction is on the order of 100 nM under these conditions. The non-specific binding of DNA that is observed at higher [ParB] might be accompanied by the formation of large nucleoprotein networks, given that these complexes are very poorly resolved by EMSA. This non-specific binding also appears to be cooperative as no intermediary species are observed and quantification of the gels reveals a highly sigmoidal pattern for the formation of this species (Figure 4.3, lower panels). We obtained values of $K_A = 230 \pm 7$ μM for *parS* substrates and $K_A = 222 \pm 8$ μM for non-specific substrates. The fact that we did not see a substantial effect of *parS* on the formation of this higher-order complex suggests that *parS* does not act as a nucleation site for ParB binding to adjacent non-specific DNA.

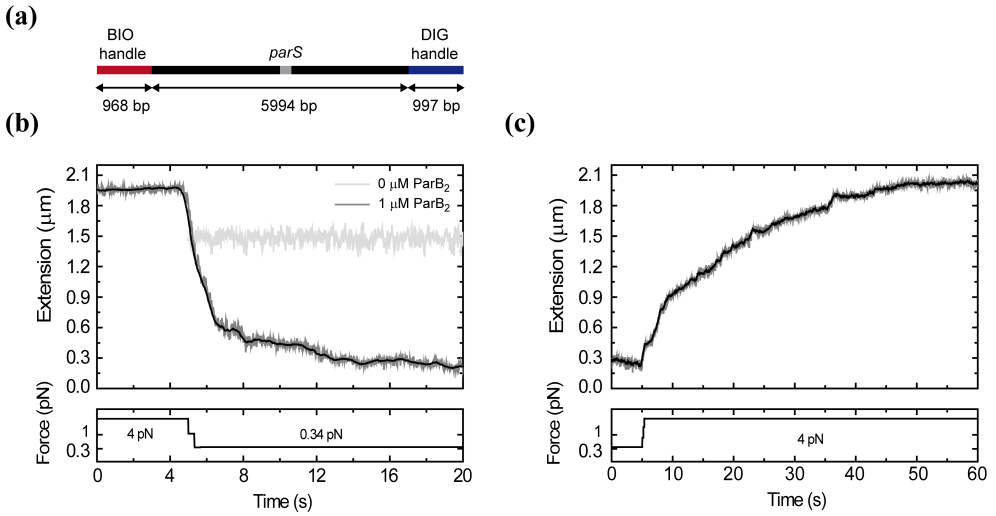

Figure 4.4: Condensation of DNA by ParB monitored by magnetic tweezers. a) Schematic representation of the *parS* DNA substrate. b) Condensation assay. At 4 pN stretching force, 1 µM $ParB_2$ was injected in the fluid cell and incubated for two minutes. Following incubation, the force was reduced to 0.34 pN. In absence of protein this leads to the change in extension represented in the light grey trace. However, in the presence of ParB we observed a progressive decrease of the extension until reaching a final extension near the surface. c) Condensed DNA molecules can be decondensed exerting high forces. Decondensation is monotonic and slower than condensation. Raw data was acquired at 60 Hz (red) and filtered down to 2.4 Hz (black).

To quantitatively study the binding of ParB we developed a MT assay to follow the binding of ParB to both *parS*-containing and non-specific DNA in real-time. In our assays we worked under conditions suitable for specific binding, and in which the forces associated with condensation and decondensation can be rigorously investigated. In these experiments, a DNA substrate of 6 kb (Figure 4.4a) containing a single *parS* site at its centre is attached to a coverslip at one end via DIG:anti-DIG interactions, and to a magnetic bead at the other end by biotin:streptavidin interactions. The bead can then be manipulated by raising, lowering and twisting a pair of magnets above it, and its height is tracked by the diffraction patterns it projects under LED light. This allows precise control of the force and twist applied to multiple DNA molecules and their heights can be monitored simultaneously. In the absence of ParB, the height of the bead, and therefore the extension of the DNA, is simply dependent on the force applied: a drop in force from 4 to 0.34 pN results in a small drop in extension from 1.9 µm to 1.5 µm, where the theoretical length of the fully extended substrate is ≈2 µm (Figure 4.4b). At a constant force of 4 pN, addition of 1 µM $ParB_2$ had no effect on the extension of the DNA substrate. However when the force was subsequently dropped to 0.34 pN, we observed a decrease in extension that was substantially greater than that attributable to the decrease in force alone. The DNA shortened from the expected 1.5 µm to an average extension of 0.2 µm, consistent with the recently reported condensation activity of ParB [Graham et al.,

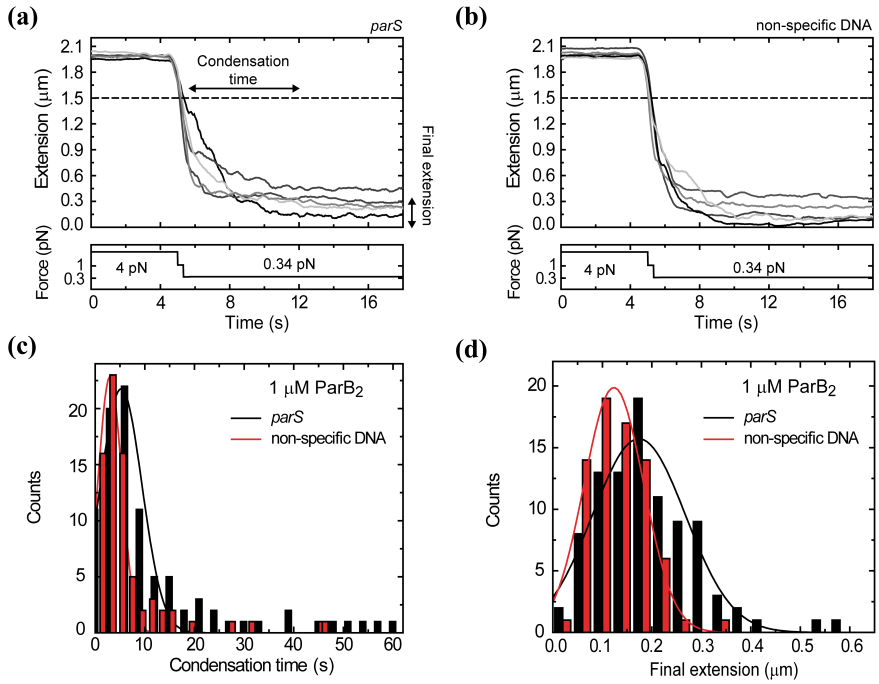

Figure 4.5: ParB-mediated DNA condensation parameters. a) Representative condensation curves for a *parS*-containing DNA substrate. b) Representative condensation curves for a non-specific DNA substrate. The dashed line represents the expected extension for a force of 0.34 pN in the absence of protein. The presence of *parS* makes no apparent difference to the process of ParB-dependent condensation of DNA. Smoothed condensation traces at 1 μM $ParB_2$. c) Distribution of condensation times for *parS* (black) and non-specific (red) DNA substrates. d) Distribution of final extensions after condensation for *parS* (black) and non-specific (red) DNA substrates.

2014]. Stretching the DNA molecules at 4 pN results in progressive increase of the extension of the tether (Figure 4.4c). Decondensation is monotonic, smooth and requires longer times than condensation. This could reflect the generation of additional condensed structures not observed during the condensation process. However, we cannot discard the effect of the interaction between bound proteins with the surface.

Similar results were obtained for a substrate lacking on the *parS* site, suggesting that *parS* binding is not required for this activity of ParB (Figure 4.5). In Figures 4.5a and 4.5b examples of condensation traces for the *parS* and for a non-specific DNA substrate are represented. By combining the observations of multiple beads, we calculated the average time taken for condensation and the average extension after the condensation for both substrates. We found an average condensation time of 5 $\pm$ 4 s and an average final extension of 0.18 $\pm$ 0.08 μm (Figures 4.5c and 4.5d) for the *parS* containing substates. We observed a slightly lower condensation time (3 $\pm$ 2 s) and final extension (0.12 $\pm$ 0.06 μm) for the non-specific DNA substrate, but these differences were within the standard deviation of the data distributions. The final extension and condensation time displayed large variability, with no observable

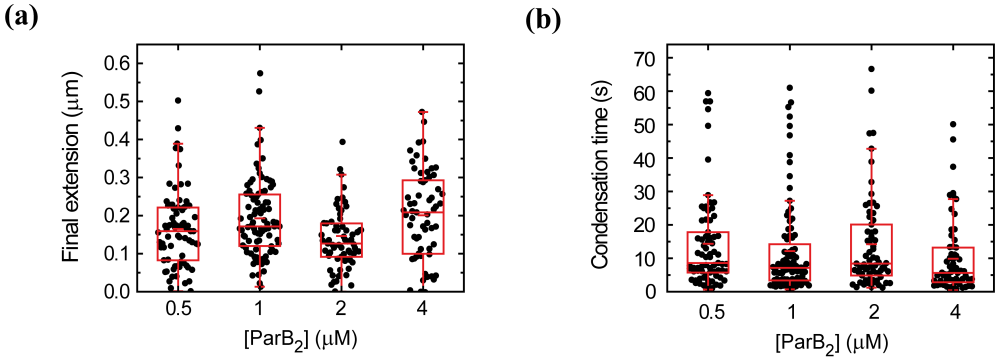

Figure 4.6: Condensation time and final extension after condensation are independent of the ParB concentration. a) Final extension measured for different protein concentrations. b) Condensation times measured for different protein concentrations. The data were obtained using the *parS* DNA substrate at $F = 0.34$ pN.

trend at the varying concentrations tested, at least at the studied force (0.34 pN) (Figure 4.6). This indicates saturating conditions of concentration, for $[\mathrm{ParB}_2] > 0.5$ μM and thus is expected full coverage of the DNA. This point will be discussed later.

As indicated previously, the condensation process was fully reversible by returning the applied force to 4 pN (Figure 4.7a). The decondensation process is highly heterogeneous and presents periods of pauses of random duration and runs of variable velocity. This results in large differences among decondensation traces for different molecules. Considering this experimental variability we did not observe noteworthy and distinct features during decondensation of *parS* substrates when compared with non-specific substrates (Figure 4.7a and b). This again contrasts with the hypothesis of the *parS* sequence as an organization site for non-specific binding and condensation.

Condensation is also reversible by the introduction of either *parS* or non-specific competitor oligonucleotides to the flow cell (Figure 4.8). In this assay, the *parS* DNA substrate is condensed as previously described by decreasing the force from 4 pN to 0.3 pN. Immediately after the reduction of force, an excess of a free competitor DNA (5 μM) containing the *parS* sequence is injected in the flow cell (Figure 4.8a). The lenght of the DNA fragment (50 bp) ensures that just a single dimer per competitor DNA molecule can be bound considering the 30 bp footprint of the ParB dimer [Murray et al., 2006]. The exchange of buffer results in progressive decondensation, characterized by a stepwise increase of the extension. Decondensation can be complete as revealed by the full recovery of the expected extension for the applied force (0.34 pN). The complete inhibition of the condensation activity is confirmed by reducing the applied force to 0 pN followed by stretching to 0.34 pN. We do not observe the characteristic pattern of force-induced decondensation neither residual condensation since full extension is recovered. The same experiments are conducted by injection of non-specific competitor DNA molecules, observing similar decondensation traces (Figure 4.8b). Nonetheless, decondensation proceeds at a slower rate

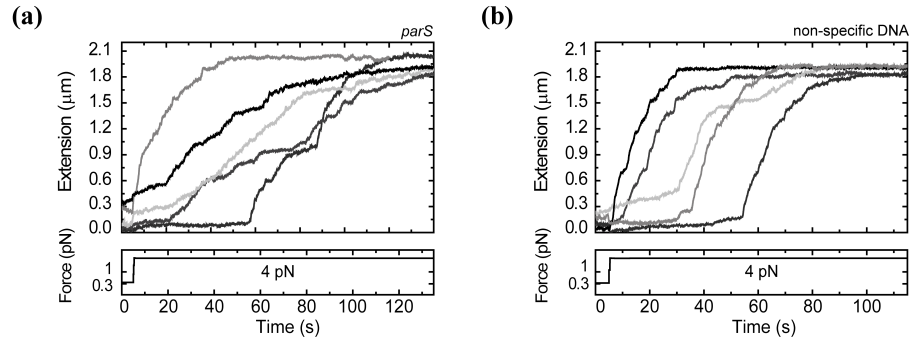

Figure 4.7: ParB-dependent condensation of DNA is reversible by force. a) Representative decondensation curves for a *parS*-containing substrate and b) non-specific DNA. No significant differences are observed between substrates considering the variability among traces.

compared with decondensation induced by *parS* competitor. In addition, full extension for the applied force is usually not recovered. The higher efficiency of the *parS* competitor in decondensing ParB-dependent condensation reflects the preferential binding of ParB to its specific sequence. We perform additional experiments where competitor DNAs are co-incubated with ParB for a few minutes before being injected in the flow cell. In agreement with decondensation experiments, co-incubation of ParB with competitor DNAs results in a reduction or even in a complete inhibition of the condensation. At a concentration of 0.5 μM the *parS* competitor completely inhibits ParB condensation at all the forces tested (Figure 4.8c). Non-specific DNA competitor inhibits ParB condensation at 0.34 pN but it was necessary to increase the concentration up to 1.5 μM and residual condensation is still present at lower forces (Figure 4.8d). Overall, competition experiments reflects the ability of the protein to discriminate between specific and non-specific DNAs at our single-molecule conditions and discards DNA condensation as a consequence of protein aggregation.

There are significant differences observed when comparing decondensation curves produced by competitor oligonucleotides versus those produced by application of high force. The former are generally much faster (seconds versus minutes) and are characterized by large steps, whereas the latter are quite variable, lasting from seconds to minutes, and feature many small extension steps. This suggests a substantially different mechanism of decondensation in each case and this will be discussed below. In either case, the total number of steps required to return to full extension varied substantially between individual experiments, which again suggests that the condensed state is not highly ordered.

The process of condensation was further analysed by varying force at a fixed ParB concentration (Figure 4.9a). The force-extension curves show that the condensation force at 1 μM $ParB_2$ is around 1 pN, and the errors reflect the variability between different molecules employed in this study (N = 15). We did not observe any effect on the contour length L of the DNA when incubated with ParB. This suggests absence of ParB-mediated bending of the DNA. The presence or absence of a single *parS*

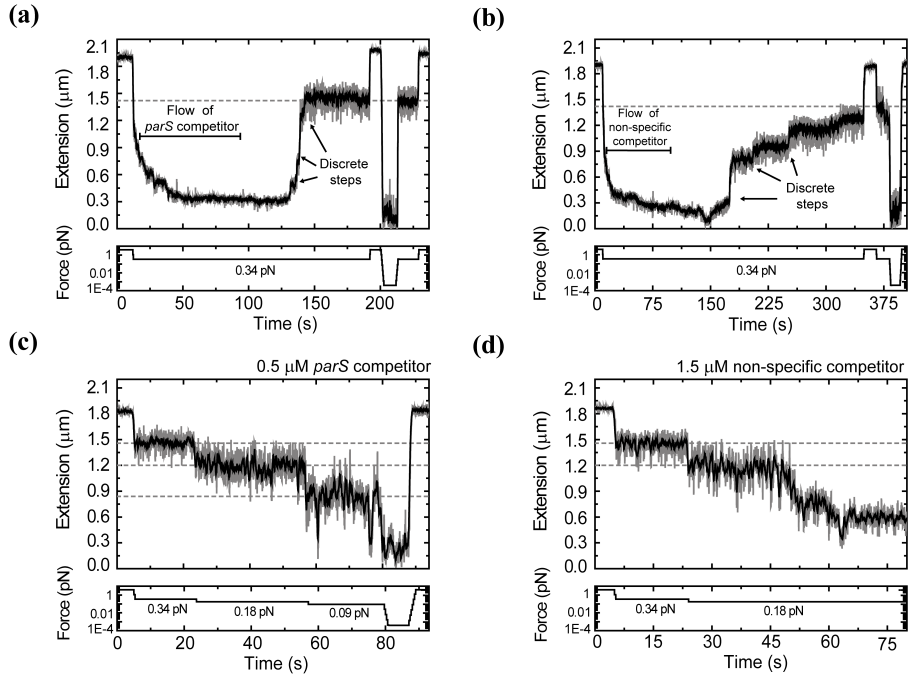

Figure 4.8: ParB-dependent condensation of DNA is reversible, and inhibited, by addition of free DNA molecules. a) Decondensation of DNA by *parS* competitor DNA. Decondensation stopped at the extension expected for 0.34 pN applied force in the absence of protein (indicated by the grey dashed line). b) Decondensation of *parS* DNA substrate by non-specific competitor DNA. c) Condensation was inhibited by the presence of *parS* competitor DNA from the outset of the experiment. d) Condensation was also inhibited by the presence of non-specific competitor DNA, although a larger concentration was required to prevent condensation in comparison to specific *parS* competitor. The expected extension at the indicated forces in the absence of protein is shown by the grey dashed lines.

has no effect on the apparent condensation force induced by ParB and both force-extension curves overlap. We then normalized the curves by their contour length for comparison. At the rather low forces employed in this study, overlapping of *parS* and non-specific DNA force-extension curves in the presence of ParB does not favour a hypothetical scenario where condensation is a consequence of local reorganization of the DNA structure dependent on protein spreading from *parS* sequences.

In order to characterize in detail the transition force between stretched and condensed DNA we set a change of extension as threshold of the occurrence of condensation events (see Materials and Methods). We termed condensation force to the first force at which the threshold in extension is exceeded. This allowed us to determine the maximum force that prevented condensation at a given [ParB] (Figure 4.9b). The apparent condensation force exerted by ParB is linear up to concentrations of 2.5 μM $ParB_2$, and saturates at ∼5 μM $ParB_2$ with a maximum value of ∼2.1 pN. This concentration-dependent result apparently contrasts with the previously found independence of the concentration for the final extension and the condensation time. Nonetheless, in the calculation of the condensation force, the threshold

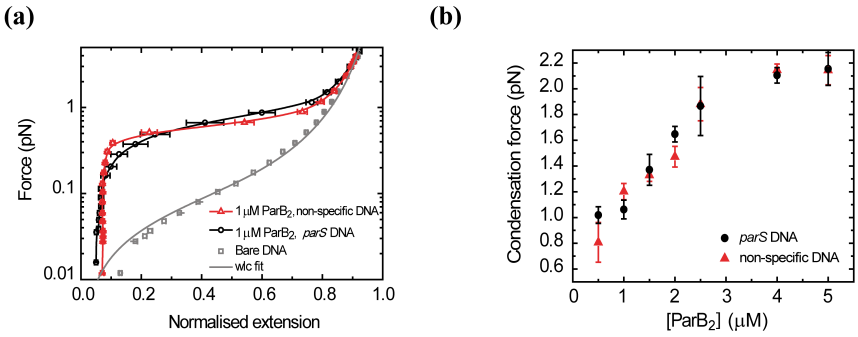

Figure 4.9: The *parS* sequence does not affect the range of permissive forces for condensation. a) Mean force-extension curve of DNA molecules in the presence 1 μM $ParB_2$ for the *parS* substrate (black) and the non-specific substrate (red). The solid red and black lines are included as a guide for the eye. Grey data corresponds to the control experiment in the absence of protein. Solid grey line is a fit to the worm-like chain model. Errors are s.e.m. of measurements on different molecules. b) Condensation force dependency on ParB concentration. A maximum condensation force of 2.1 pN was measured at saturating protein concentration for both *parS* and non-specific DNA substrates. Errors are the standard deviation of measurements on different molecules.

was set to discriminate subtle changes in extension and thus reflects the first steps of the condensation process. This indicates that while the degree of coverage affects the range of permissive forces, the global efficiency of the condensation process is not significantly affected at low forces, at least for the concentrations explored in this work.

The maximum condensation force found is similar to those reported for Fis and HU, other bacterial proteins with a role in bacterial chromosome organisation [Skoko et al., 2005; Xiao et al., 2010]. This is consistent with the idea that condensation is induced via ParB-ParB interactions, and that these are allowed by DNA fluctuations which are limited by the applied force. For instance, a high force of 2 pN restricts the extension fluctuation of our DNA to 30 nm, and therefore a very high coverage of the DNA (produced by a high concentration of proteins) is required for condensation. The fact that we observe no condensation at forces over 2.1 pN indicates either that the proteins are not close enough to interact, or simply that the protein:DNA and/or protein:protein interactions involved in condensation cannot withstand forces over 2.1 pN. Remarkably, we found the same trend and the same saturation force in the *parS*-containing substrate and in the non-specific DNA substrate. This contrasts with the current model for the *parS* sequence as nucleation site, since the generation of a densely coated and organized region spreading from the *parS* site should strongly affect the range of permissive forces.

4.3.2 ParB does not form ordered structures on the DNA

In order to probe the structure of the condensed nucleoprotein complex, we analyzed the relation between the initial extension and the final extension after condensation.

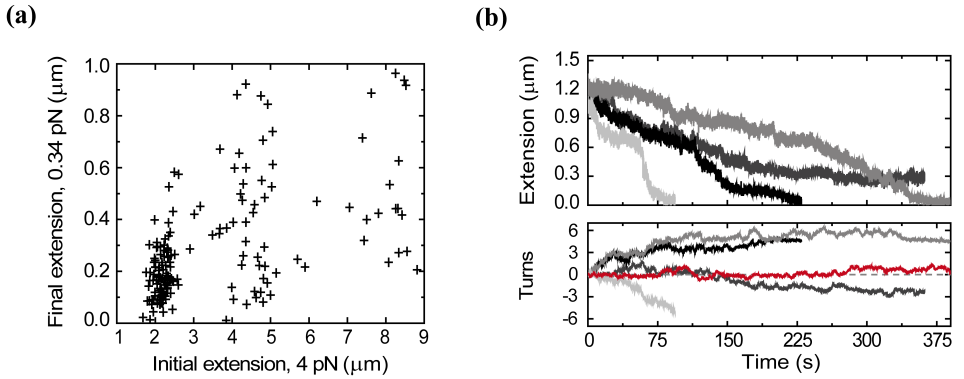

Figure 4.10: ParB condensation is not dependent on ordered structures. a) Scatter plot of initial and final extensions for substrates of different contour lengths. All of the data shown were obtained from condensation curves of 1 μM $ParB_2$ at 0.34 pN. b) ParB-dependent condensation assay using freely orbiting magnetic tweezers and torsionally-constrained DNA molecules. The turns made by the bead during condensation (F = 0.45 pN) are both clockwise and anti-clockwise, small in number (<6 at most), and uncorrelated with changes to extension. Different molecules represented in shades of gray. Turns for a bare DNA shown in red.

We performed these experiments using substrates with different contour lengths. Importantly, the final extension is not proportional to the length of the DNA substrate used and was highly variable for a given length (Figure 4.10a). This is inconsistent with a model in which condensation is due to the formation of an ordered species such as a continuous nucleoprotein filament, and where condensation is dependent on local modification of the DNA structure upon protein binding. Freely-orbiting magnetic tweezers experiments that are presented below also support this conclusion.

We used freely-orbiting magnetic tweezers (FOMT) to provide a measurement of the twist induced by the binding and condensation of ParB to torsionally-constrained (TC) *parS* DNA molecules. Prior to any attempt to measure a signal induced by the binding of a protein in the FOMT, we checked that the DNA molecules were TC using the standard MT. We also checked that TC DNA molecules were condensed by ParB with no detectable differences compared to non-coilable molecules. FOMT provided no evidence for concerted twisting of the DNA as condensation occurs since the bead turns are both positive and negative, small in number and do not display correlation between the velocity of rotation and the velocity of condensation (Figure 4.10b). This behavior of ParB is in stark contrast to that displayed by an ordered filament-forming protein such as RecA, where directional untwisting is tightly correlated with DNA extension (section 3.7.2, Figure 3.12). In fact, condensation experiments on TC molecules show no significant differences in condensation efficiency when subjected to positive or negative supercoiling as would be expected for a protein inducing supercoiling (data not shown).

All together, these experiments suggest that ParB condensation is not associated to a structural distortion of the DNA neither to DNA wrapping.

4.3.3 ParB stabilises looping interactions between distal DNA segments

To test the formation of ParB-mediated interactions formed between distal DNA segments *in trans*, we next performed DNA braiding experiments (Figure 4.11a). In these experiments, a single bead holding two DNA molecules was held at a force non-permissive to condensation (4 pN). Rotation of the magnets in either direction causes the two DNA duplexes to crossover, which is observed as a decrease in the height of the bead.

As expected, in the absence of ParB, the removal of this twist unbraids the DNA strands and restores the height of the bead since no-interaction occurs between the two DNA molecules (Figure 4.11b). However, in the presence of 1 μM $ParB_2$, unbraiding is prevented and the bead does not return to its original height upon untwisting (Figure 4.11c). This suggests that the two DNA duplexes are held together by ParB. The handedness of the twist had no detectable effect on the stabilisation of the braid by ParB (data not shown). The tracking of several of these doubly-tethered beads simultaneously showed that this stabilisation phenomenon was highly reproducible. The introduction of a *parS*-containing competitor oligonucleotide restored the bead to its original height after untwisting, indicating that braiding-stabilisation is dependent on ParB binding to DNA (Figure 4.11d).

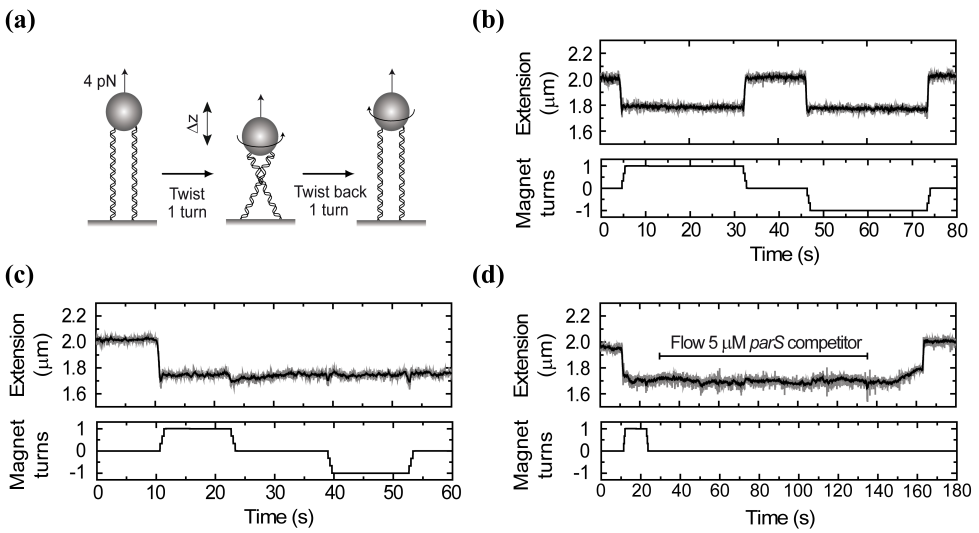

Figure 4.11: ParB stabilizatation of bridging interactions *in trans*. a) Illustration of the experiment to braid DNA segments *in trans*. The application of one turn (clockwise or anticlockwise) to doubly tethered beads promotes the cross-over of both DNAs, leading to a change of the extension (Δz). Subsequent untwisting to zero rotation immediately recovers the original extension. b) Time trace of an experiment on bare DNA. c) In the presence of ParB the cross-over is stabilised, and the extension does not recover after untwisting to zero rotation or even when one additional turn in the direction opposite to the cross-over was applied. d) Injection of 5 μM *parS* DNA competitor oligonucleotide promotes the recovery of the full extension following ParB-mediated stabilisation of a braid.

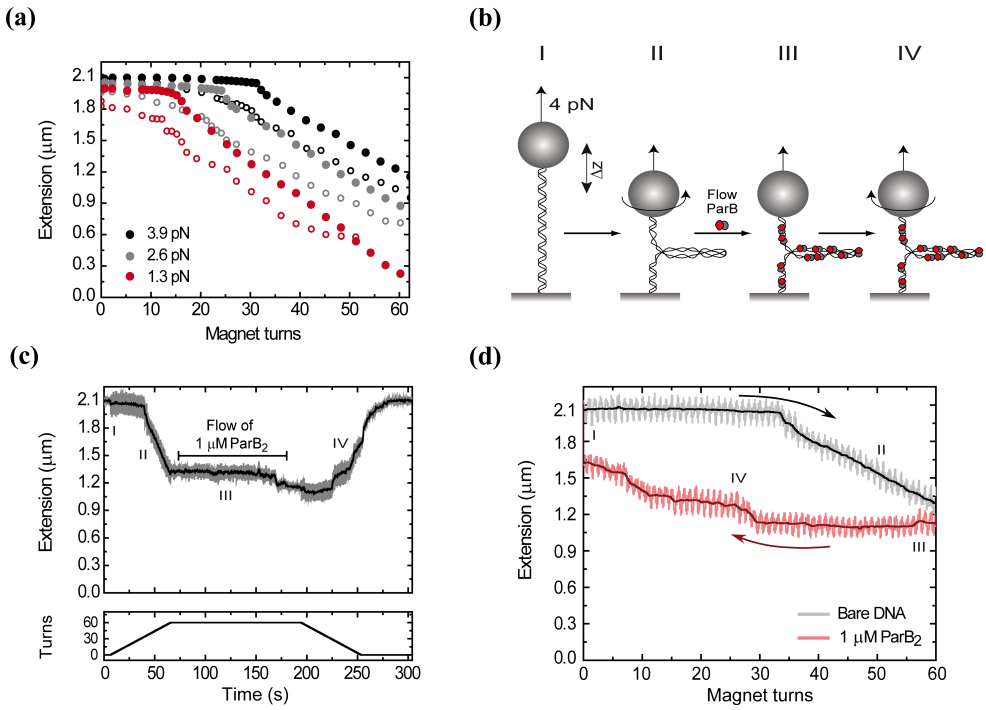

Figure 4.12: ParB stabilisation of bridging interactions *in cis*. a) Rotation-extension curves at 3.9, 2.6 and 1.3 pN for bare DNA (filled symbols) and in the presence of 1 µM $ParB_2$ (open symbols). ParB-coated DNA molecules shows a reduction in the number of turns necessary to buckle the DNA. b) Cartoon of the experiment to bridge DNA segments *in cis*. Single torsionally-constrained DNA molecules (I) are positively supercoiled at 4 pN force by applying 60 turns (II). Then, 1 µM $ParB_2$ is injected into the fluid cell (III). After full buffer exchange, all of the turns are released (IV). c) Dynamics of the plectoneme stabilisation experiment highlighting the injection of ParB. d) DNA extension is displayed as a function of turns to highlight the hysteresis observed due to bridging of different regions of supercoiled DNA after introduction of ParB. The numbers indicate the different stages of the experiment as per the cartoon.

In addition, we tested for the role of ParB in DNA supercoiling. First, we performed experiments where ParB is incubated on TC DNA molecules at non-permisive forces (Figure 4.12a). Then, we introduced positive turns and this results in a change of the twist of the DNA molecule with no-observable decrease of extension until the DNA buckles. Following addition of turns induces the formation of plectonemes, what can be monitored from the decrease of the extension of the tether with the applied turns. We find a decrease in the number of turns required to buckle the DNA in the presence of ParB when compared with bare DNA. The reduction of turns to buckle the DNA in the presence of ParB is force dependent, and is on average -8.5 $\pm$ 3.7 turns for 3.9 pN and -8 $\pm$ 3.6 tuns at 2.6 pN. This reduction of the buckling number is engaged with a smoother transition between stretched and plectonemic DNA. However, there is no alteration of the relation between the change of

extension with the applied turns (Δl) after buckling. This indicates that while ParB facilitates the formation of the first supercoil by favouring the formation of small plectonemes, it has a minor effect on the formation of subsequent plectonemes.

A simple mechanical model describe this slope as $\Delta l = \pi\sqrt{2k_BT\xi F^{-1}}$, where ξ is the persistence length and F the applied force [Strick et al., 2003]. Since $\Delta l_{\mathrm{DNA}} \approx \Delta l_{\mathrm{ParB}}$ we can conclude that the persistence lenght of the DNA is not significantly affected by ParB binding. It has been described that protein-induced bending results in a change of the apparent persistence length of the molecule [Kulic et al., 2005]. Therefore, our results again support the idea that ParB binding does not disturb the structure of the DNA during binding and condensation.

We performed additional experiments to test for the ability of ParB to stabilise remote DNA regions of a single DNA molecule. For this we selected TC molecules and applied positive turns to induce plectoneme extrusion while stretching at non-permissive forces, as previously described. Then, we injected one cell volume of 1 μM $ParB_2$ and after buffer exchange, we released the applied turns. A scheme of the experiment is represented in Figure 4.12b. We found that ParB was capable of stabilising the supercoils introduced, determined from the reduction of the extension for a given number of applied turns (Figures 4.12c and 4.12d). This is consistent with a model of ParB bound along the length of the DNA and stabilising the plectonemes of the supercoiled DNA by protein-protein interactions *in cis*.

Together, these data indicate that ParB is capable of bridging between separate DNA duplexes *in trans*, as well as between different segments of the same DNA molecule *in cis*.

4.4 Discussion

ParB forms large nucleoprotein complexes that spread for several kilobase pairs around individual *parS* sites *in vivo*. These act as markers for active segregation and as loading sites for SMC complexes. The mechanisms for ParB spreading, recruitment of SMC complexes, and for the ensuing segregation and condensation of the bacterial chromosome are not fully understood. In this work, we studied the binding of purified ParB protein to DNA *in vitro*. Electrophoretic mobility shift assays revealed a clear specificity for binding to the *parS* sequence. At elevated ParB concentrations, there is a non-specific binding of DNA which results in nucleoprotein complexes that are almost unable to migrate into native gels. This assay suggested that the non-specific binding is strongly co-operative (Hill co-efficient ∼3-5) on the substrates used here and, importantly, there is no substantial effect of *parS* sequence on the binding isotherms. These findings were corroborated in single molecule MT experiments. We carefully and deeply explored the role of the *parS* sequence in DNA condensation at variable forces, obtaining similar values for all the observables analyzed in *parS* and non-specific DNA substrates.

Recent work using single molecule visualisation and complementary modelling of ParB binding to DNA has shown that ParB is capable of condensing DNA, and suggests that this activity is the result of bridging interactions between many non-

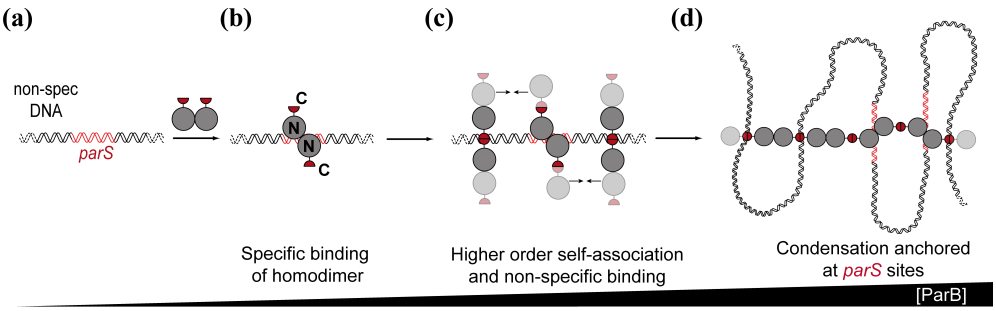

Figure 4.13: Specific and non-specific binding of DNA by ParB: speculative model for spreading at *parS* sites. a) A region of a DNA molecule containing a specific binding site is shown b) Specific binding. At low concentrations, ParB binds to *parS* sequences via the central helix-turn-helix motifs to form a $ParB_2$:DNA complex. c) Non-specific DNA binding. Elevated concentrations of ParB allow co-operative non-specific binding via a second (hypothetical) DNA binding domain. The continued self-association of ParB (indicated with arrows) via at least two interfaces subsequently leads to formation of higher order networks and DNA condensation d) The condensed nucleoprotein network contains both specific and non-specific DNA binding sites that trap loops of DNA that are anchored around *parS* (if *parS* is present). For simplicity, the specific binding sites for most of the ParB dimers are shown unoccupied. Such structures might bridge larger distances, including between distant *parS* loci, through the sharing of segments of DNA, or via additional protein:protein interactions.

specifically bound ParB molecules [Graham et al., 2014; Broedersz et al., 2014]. Consistent with these data, we found using magnetic tweezers that ParB was able to condense DNA against a maximum applied force of 2.1 pN. This is a fairly small force that could easily be applied by the many cellular motor proteins that act on DNA. An attractive idea, therefore, is that ParB-mediated condensation may be modulated by other processes that remodel bacterial chromatin via stretching of the chromosome. Condensation was dependent on the ParB-DNA interaction as it was both inhibited by, and fully reversible through, the addition of free DNA. Condensation was also reversible by applying a large force, although these decondensation traces were markedly different from those obtained by the addition of competitor DNA. The different behaviour in each case is consistent with a model in which multiple ParB protomers stabilise large loops of DNA in the condensed state. In that case, decondensation by force would selectively disrupt the ParB-DNA interactions at the edges of such structures, leading to DNA extension in small steps as the loops become progressively smaller. In contrast, addition of competitor DNA would randomly disrupt the interactions, such that when the critical interactions at the edge of condensed loops are eventually eliminated, they are likely to release larger amounts of DNA, leading to extension in bigger steps as observed.

A model in which ParB bridges distant DNA segments to form loops is further supported and extended by our observations that ParB can stabilise writhe in supercoiled DNA molecules, and can hold together braids between two DNA molecules. Importantly, these experiments also exclude the possibility that the condensation ac-

tivity we observe is due to interactions of ParB with the surface of the flow cells used in our apparatus. Additionally, we found that the condensed DNA structure formed by ParB was not well ordered. The final extension was highly variable and only weakly correlated with the initial DNA length. Moreover, freely orbiting magnetic tweezers experiments showed that condensation is not coupled to DNA twisting in a concerted fashion, as is the case for classical nucleoprotein filament forming proteins such as RecA. We conclude that *Bacillus* ParB (a Type Ia centromere binding protein) does not form structures that are similar to the ordered large *superhelix* complexes that have been suggested for some Type II systems such as pSK41 ParR [Schumacher, 2008; Schumacher et al., 2007].

A working model for ParB-dependent condensation of DNA is shown in Figure 4.13. It is somewhat similar to recently published models [Graham et al., 2014; Broedersz et al., 2014], which explained how large regions of DNA around *parS* could be condensed by relatively few ParB protomers and rationalised the effects of roadblocks on ParB distribution in terms of a bridging model (as opposed to the, perhaps more intuitive, lateral spreading or filament model shown in Figure 4.2). That work envisioned that the ParB protein contained a single DNA binding site, and therefore assumed that the non-specific DNA binding required for looping and spreading must be competitive with specific binding. In this single DNA binding site scenario, given the very high concentration of non-specific versus specific sites *in vivo*, localisation of the ParB nucleoprotein complexes at *parS* would require an exceptionally high degree of specificity of binding to *parS* sequences. However, this is not well supported by the available data for the behaviour of ParB *in vitro* [Lin and Grossman, 1998; Murray et al., 2006]. In fact, the work presented here shows no apparent difference between specific and non-specific binding affinity at all. Therefore, an alternative idea is that the non-specific binding we observe may occur at a dedicated and structurally distinct DNA binding locus, as is the case for other type I centromere binding proteins [Schumacher, 2008]. Such a binding site might belie the potentially high specificity for binding *parS*. In such a scenario, large protein networks formed by ParB protein through its self-association interfaces, would contain both highly-specific *parS* binding sites, as well as additional non-specific sites (Figures 4.13c and 4.13d). DNA condensation might require the co-operative binding of multiple ParB proteins present at relatively high concentration and could occur anywhere on the chromosome. However, in a situation where the total protein is limited relative to DNA binding sites, as is the case *in vivo*, these networks would be preferentially anchored around *parS* by the additional binding energy available at the specific binding locus. These ideas will be tested and developed further in the following chapter.

References

Autret, S., Nair, R., and Errington, J. Genetic analysis of the chromosome segregation protein Spo0J of bacillus subtilis: evidence for separate domains involved in DNA binding and interactions with Soj protein. *Mol. Microbiol.*, 41(3):743–55, 2001.

Bartosik, A. A., Mierzejewska, J., Thomas, C. M., and Jagura-Burdzy, G. ParB deficiency in Pseudomonas aeruginosa destabilizes the partner protein ParA and affects a variety of physiological parameters. *Microbiology*, 155(4):1080–92, 2009.

Bouchiat, C., Wang, M. D., Allemand, J., Strick, T., Block, S. M., and Croquette, V. Estimating the persistence length of a worm-like chain molecule from force-extension measurements. *Biophys. J.*, 76 (1):409–413, 1999.

Breier, A. M. and Grossman, A. D. Whole-genome analysis of the chromosome partitioning and sporulation protein Spo0J (ParB) reveals spreading and origin-distal sites on the bacillus subtilis chromosome. *Mol. Microbiol.*, 64(3): 703–18, 2007.

Britton, R. A., C., L. D., and Grossman, A. D. Characterization of a prokaryotic SMC protein involved in chromosome partitioning. *Genes Dev.*, 12(9):1254–9, 1998.

Broedersz, C. P., Wang, X., Meir, Y., Loparo, J. J., Rudner, D. Z., and Wingreen, N. S. Condensation and localization of the partitioning protein ParB on the bacterial chromosome. *Proc. Natl. Acad. Sci. U.S.A.*, 111(24): 8809–14, 2014.

Dame, R. T. The role of nucleoid-associated proteins in the organization and compaction of bacterial chromatin. *Mol. Microbiol.*, 56(4):858–70, 2005.

Fili, N., Mashanov, G. I., Toseland, C. P., Batters, C., Wallace, M. I., Yeeles, J. T., Dillingham, M. S., Webb, M. R., and Molloy, J. E. Visualizing helicases unwinding DNA at the single molecule level. *Nucleic Acids Res.*, 38(13):4448–57, 2010.

Gerdes, K., Møller-Jensen, J., and Bugge Jensen, R. Review plasmid and chromosome partitioning: surprises from phylogeny. *Mol. Microbiol.*, 37(3):455–466, 2000.

Gerdes, K., Howard, M., and Szardenings, F. Pushing and pulling in prokaryotic DNA segregation. *Cell*, 141(6):927–942, 2010.

Graham, T. G., Wang, X., Song, D., Etson, C. M., van Oijen, A. M., Rudner, D. Z., and Loparo, J. J. ParB spreading requires DNA bridging. *Genes Dev.*, 28 (11):1228–38, 2014.

Griswold, A. Genome packaging in prokaryotes: the circular chromosome of *E. coli*. *Nature Education*, 1(1), 2008.

Gruber, S. and Errington, J. Recruitment of condensin to replication origin regions by ParB/Spo0J promotes chromosome segregation in B. subtilis. *Cell*, 137(4):685–696, 2009.

Hirano, T. At the heart of the chromosome: SMC proteins in action. *Nat. Rev. Mol. Cell. Biol.*, 7(8):311–22, 2006.

Hirano, T. SMC proteins and chromosome mechanics: from bacteria to humans. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.*, 360(1455):507–14, 2009.

Kulic, I. M., Mohrbach, H., Lobaskin, V., Thaokar, R., and H., S. Apparent persistence length renormalization of bent DNA. *Phys. Rev. E. Stat. Nonlin. Soft. Matter. Phys.*, 72(4 Pt. 1):041905, 2005.

Leonard, T. A., Butler, P. J., and Löwe, J. Structural analysis of the chromosome segregation protein Spo0J from Thermus thermophilus. *Mol. Microbiol.*, 53 (2):419–432, 2004.

Lewis, P. J. and Errington, J. Direct evidence for active segregation of oriC regions of the Bacillus subtilis chromosome and co-localization with the Spo0J partitioning protein. *Mol. Microbiol.*, 25 (5):945–954, 1997.

Lin, D. C. and Grossman, A. D. Identification and characterization of a bacterial chromosome partitioning site. *Cell*, 92(5):675–685, 1998.

Mierzejewska, J. and Jagura-Burdzy, G. Prokaryotic parA-parB-parS system links bacterial chromosome segregation with the cell cycle. *Plasmid*, 67(1):1–14, 2012.

Mohl, D. A., Easter, J. J., and Gober, J. W. The chromosome partitioning protein, ParB, is required for cytokinesis in Caulobacter crescentus. *Mol. Microbiol.*, 42(3):1–14, 2001.

Murray, H., Ferreira, H., and Errington, J. The bacterial chromosome segregation protein Spo0J spreads along dna from parS nucleation sites. *Mol. Microbiol.*, 61(5):1352–61, 2006.

Oliva, M. A. Segrosome complex formation during DNA trafficking in bacterial cell division. *Front. Mol. Biosci.*, 3(51), 2016.

Phillips, R., Kondev, J., and Theriot, J. *Physical Biology of the Cell.* Garland Science, 1st edition, 2009.

Rodionov, O. and Yarmolinsky, M. Plasmid partitioning and the spreading of P1 partition protein ParB. *Biochem. J.*, 52(4):1215–23, 2004.

Schumacher, M. A. Structural biology of plasmid partition: uncovering the molecular mechanisms of DNA segregation. *Biochem. J.*, 412(1):1–18, 2008.

Schumacher, M. A. Bacterial plasmid partition machinery: a minimalist approach to survival. *Curr Opin Struct Biol.*, 22(1):72–9, 2011.

Schumacher, M. A., Glover, T. C., Brzoska, A. J., Jensen, S. O., Dunham, T. D., Skurray, R. A., and Firth, N. Segrosome structure revealed by a complex of ParR with centromere DNA. *Nature*, 450(7173):1268–71, 2007.

Skoko, D., Yan, J., Johnson, R. C., and Marko, J. F. Low-force DNA condensation and discontinuous high-force decondensation reveal a loop-stabilizing function of the protein Fis. *Phys. Rev. Lett.*, 95(20):208101, 2005.

Strick, T., Dessinges, M. N., Charvin, G., Dekker, N. H., Allemand, J. F., Bensimon, D., and Croquette, V. Stretching

of macromolecules and proteins. *Rep. Prog. Phys.*, 66:1–45, 2003.

Sullivan, N. L., Marquis, K. A., and Rudner, D. Z. Recruitment of SMC by ParB-parS organizes the origin region and promotes efficient chromosome segregation. *Cell*, 137(4):697–707, 2009.

Taylor, J. A., Pastrana, C. L., Butterer, A., Pernstich, C., Gwynn, E. J., Sobott, F., Moreno-Herrero, F., and Dillingham, M. S. Specific and non-specific interactions of ParB with DNA: implications for chromosome segregation. *Nucleic Acids Res.*, 43(2):719–731, 2015.

Thanbichler, M., Viollier, P. H., and L., S. The structure and function of the bacterial chromosome. *Curr. Opin. Gene.t Dev.*, 15(2):153–62, 2005.

Wang, X., Montero Llopis, P., and Rudner, D. Z. Organization and segregation of bacterial chromosomes. *Nat. Rev. Genet.*, 14(3), 2013.

Xiao, B., C., J. R., and Marko, J. F. Modulation of HUŰDNA interactions by salt concentration and applied force. *Nucleic Acids Res.*, 38(18):6176Ű85, 2010.

Yamaichi, Y. and Niki, H. Active segregation by the Bacillus subtilis partitioning system in Escherichia coli. *Proc. Natl. Acad. Sci. U.S.A.*, 97(26):14656–61, 2000.

Yu, A. C., Loo, J. F. C., Yu, Samuel Kong, S. K., and Chan, T.-F. Monitoring bacterial growth using tunable resistive pulse sensing with a pore-based technique. *Appl. Microbiol. Biotechnol.*, 98(2):855–862, 2013.

5

The implications of ParB domains in DNA binding and condensation

Chapter overview. In the previous chapter a novel function of the ParB protein to condense DNA molecules was presented. A speculative model of the mechanism leading to DNA condensation was proposed based on the experimental results, but the details remained unclear. The available structural data suggested the presence of a secondary DNA binding domain located at the C-terminal domain of ParB. In this chapter, this hypothesis is tested, confirming the ability of the C-terminal domain for non-specific DNA binding, while being dispensable for specific DNA binding to *parS* sequences. Our work reveals a dual role of the C-terminal domain of ParB as both a DNA binding and dimerisation interface who works together with the N-terminal domain for spreading and bridging.

5.1 Introduction

Bacterial chromosomes are actively segregated and condensed by the ParABS system and condensin [Wang et al., 2014]. In *B. subtilis*, this machinery is physically targeted to the origin proximal region of the chromosome by eight palindromic DNA sequences called *parS* to which the ParB (Spo0J) protein binds specifically [Breier and Grossman, 2007; Lin and Grossman, 1998]. These nucleoprotein complexes act as a positional marker of the origin and earmark this region for segregation in a manner somewhat analogous to eukaryotic centromeres and their binding partners.

ParB is an unusual DNA binding protein. In addition to sequence-specific interactions with the *parS* sequence, the protein also spreads extensively around the site for about 18 kbp [Breier and Grossman, 2007; Lynch and Wang, 1995; Murray et al., 2006]. The mechanistic basis for this behaviour is not well understood and a matter of active debate. Earlier models envisioned a lateral 1D spreading around *parS* to form a filament [Murray et al., 2006; Rodionov et al., 1999], mainly because spreading can be inhibited in a polar manner by *roadblocks* placed to the side of *parS* sequences. However, ParB foci appear to contain fewer proteins than are necessary to form a filament, and single molecule analyses using direct imaging [Graham et al., 2014] and magnetic tweezers [Taylor et al., 2015] have shown that binding of DNA by ParB is accompanied by condensation. These networks were inferred to be dy-

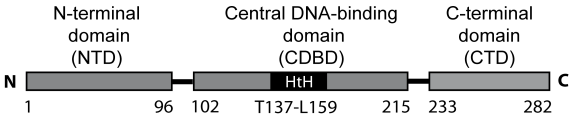

Figure 5.1: Diagram of the ParB domains. The available structural data suggest the existence of three domains. The C-terminal domain has been described as a dimerisation domain. The previously considered as N-terminal domain can be divided into i) a central DNA binding domain (CDBD) containing the helix-turn-helix (HtH) motif responsible of *parS* recognition and ii) N-terminal domain presumably involved in ParB dimerisation.

namic and poorly-ordered, consisting of many DNA loops between distantly bound ParB molecules. In cells, they are presumably anchored at *parS* sites by sequence-specific interactions but must also contain many interactions with non-specific DNA (nsDNA), as well as self-association interactions that bridge ParB protomers to form DNA loops. Modelling suggests that a combination of 1D spreading and 3D bridging interactions can explain the condensation activity and recapitulate the polar effect of roadblocks on ParB spreading [Broedersz et al., 2014]. Recently, single-molecule imaging of F-plasmid SopB led to a broadly similar model, defining ParB networks as fluid structures that localise around *parS* using a *nucleation and caging* mechanism [Sanchez et al., 2015]. Despite these recent experiments converging on DNA bridging models to explain the ParB spreading phenomenon, the mechanism underpinning this behaviour remains unresolved. In particular, the relationship between these dynamic nucleoprotein complexes and the molecular architecture of the ParB protein is unclear and this is the subject of the work presented in this chapter.

The genomically-encoded ParB proteins comprise three distinct domains (Figure 5.1). Our current understanding of their structure is limited to the N-terminal domain (NTD) which binds ParA [Bouet and Funnell, 1999; Davey and Funnell, 1997; Davis et al., 1992; Radnedge et al., 1998] and the central DNA binding domain (CDBD) which binds *parS* and possibly also nsDNA [Leonard et al., 2004; Schumacher and Funnell, 2005]. A crystal structure of *Thermus thermophilus* ParB lacking the C-terminal domain (CTD) revealed a compact dimer in which the helix-turn-helix (HtH) motifs were symmetrically arranged in a fashion that appeared suitable for binding to the palindromic *parS* sequence [Leonard et al., 2004]. Complementary analysis of the CTD by analytical ultracentrifugation suggested that it also forms a dimer, and it was argued that this protein interface might promote spreading interactions. More recently, a structure of *Helicobacter pylori* ParB, in which the protein was also truncated by removal of the CTD, showed a strikingly different conformation, where the NTD had moved away from the CDBD domain to form a tetrameric self-association interface [Chen et al., 2015]. In this structure, the CDBD was independently bound to a *parS* half site, and it was argued that the tetramerisation of the NTD could be responsible for bridging interactions between specific and nsDNA bound to the CDBD. This has yet to be tested experimentally.

In the previous chapter, we hypothesised that ParB contains a second DNA binding locus for nsDNA that functions independently of the HtH motif (Figure 4.13) [Taylor et al., 2015]. This idea was attractive to us for several reasons. Firstly, in a

single DNA binding locus model, it is not straightforward to reconcile the strict localisation of ParB networks to just a few *parS* sites and their surroundings, with the limited discrimination between specific and nsDNA binding that is observed *in vitro* ($<$ 10-fold apparent difference in affinity) [Broedersz et al., 2014; Taylor et al., 2015]. Secondly, although binding of *parS* protects the CDBD region from proteolysis, high concentrations of nsDNA afford no such protection, implying that it binds elsewhere [Taylor et al., 2015]. Thirdly, the distantly-related ParB protein from plasmid P1 provides a precedent for a second DNA binding locus in a Type I centromere binding protein, and highlights the CTD as the putative candidate region [Schumacher et al., 2007].

In this chapter, we have probed the role of the NTD, CDBD and CTD of *B. subtilis* ParB. We find that CDBD specific recognition of *parS* is not fully required for DNA condensation, while mutations at the NTD strongly deplete ParB-dependent condensation. Furthermore, we find that the CTD provides both a second nsDNA binding site and a self-association interface that promotes bridging interactions and DNA condensation.

5.2 Materials and Methods

5.2.1 Protein purification

Untagged ParB was purified by Gemma Fisher from the group of Prof. Mark S. Dillingham (University of Bristol). ParB mutations were introduced by site-directed mutagenesis (R149G) or produced synthetically (K255A-K257A, K252A-K255A-K259A) by Life Technologies (Thermo Fisher Scientific Inc., MA USA). The CTD truncations (residues 217 to 282) were produced by PCR amplication of the region of interest of the gene with primer overhangs incorporating 5' PacI and 3' XmaI restriction sites for subcloning.

5.2.2 Magnetic Tweezers

Experimental conditions
We used the home-built setup described in section 3.2.1. Data were acquired at 120 Hz and filtered down to 3 Hz for representation and analysis.

Unless specified otherwise, we used the *parS* substrate previously described in section 4.2.4. All the reactions were carried at room-temperature in a reaction buffer composed of 100 mM NaCl, 50 mM Tris-HCl pH 7.5, 4 mM $MgCl_2$, 1 mM DTT, 100 µg/ml BSA and 0.1% Tween-20.

CTD-induced decondensation
Once selected single torsionally-relaxed DNA molecules, 1 µM $ParB_2$ was incubated for 2-3 minutes and condensation was induced by decreasing the force to 0.34 pN. Immediately after, one cell-volume of reaction buffer containing 5 µM CTD_2 and 1 µM $ParB_2$ was flown at a constant flow velocity of 16 µl/min. In control experiments where only 1 µM $ParB_2$ was flown, the reaction was supplemented with a volume of storage buffer equal to that used in the CTD experiments to maintain the ionic conditions.

To quantify the degree of induced decondensation, we determined a condensation ratio, C_r, which was calculated as:

$$C_r = \frac{l_0 - l}{l_0}, \tag{5.1}$$

where, l_0 is the expected extension at 0.34 pN measured before ParB injection and l is the equilibrium extension after induced-decondensation. l, was determined from the average extension measured after the cell volume was completely exchanged (after 390 seconds) and considering a 120 data points window (2 seconds). These data were acquired at 60 Hz and filtered down to 3 Hz.

Force-extension curves and work calculation

Force-extension curves were obtained by decreasing the applied force in steps from 4 pN to $\sim$ 0.02 pN for a total measuring time of 13 minutes. This procedure is initially performed for bare DNA molecules. Then, the force is reset to 4 pN and ParB variants are flown and incubated for 2 minutes before starting the measurement of a new force-extension curve using the same magnet positions employed in absence of proteins. In every case, the force applied to each bead was calculated from the force-extension data of bare DNA molecules.

The work done during condensation ΔW can be calculated by the difference in work between the force-extension curve in the presence of ParB variants and that of bare DNA,

$$\Delta W = \int_{F_{min}}^{F_{max}} (z_{DNA}(F') - z_{ParB}(F'))dF'. \tag{5.2}$$

Numerical integration of force-extension curves was calculated using the trapezoidal rule.

5.3 Results

5.3.1 The NTD is essential for DNA condensation but *parS* recognition is dispensable

Genetic and structural analyses indicate that residue R149 is critically important for specific binding to *parS* at the HtH locus [Graham et al., 2014; Chen et al., 2015; Autret et al., 2001; Gruber and Errington, 2009]. EMSA assays with $ParB^{R149G}$ reveals a loss of *parS* binding, whilst the formation of the supershifted complex is retained, hence indicating that non-specific binding is not strongly affected [Fisher et al., 2017]. We therefore explored the effects on DNA condensation of this variant using MT. Condensation experiments showed a partial impairing of the condensation compared with wild-type ParB (Figure 5.2a). Force-extension curves (FEC) showed a highly cooperative transition (Figure 5.2b). At 1 μM $ParB_2^{R149G}$ condensation was not observed, although fluctuations of the DNA tether were consistent with minor condensation events that do not greatly affect the mean extension value measured (Figure 5.2b). Nonetheless, at a moderatelly elevated concentration (3 μM $ParB_2^{R149G}$) condensation was similar to that obtained with wild-type ParB, as indicated by the

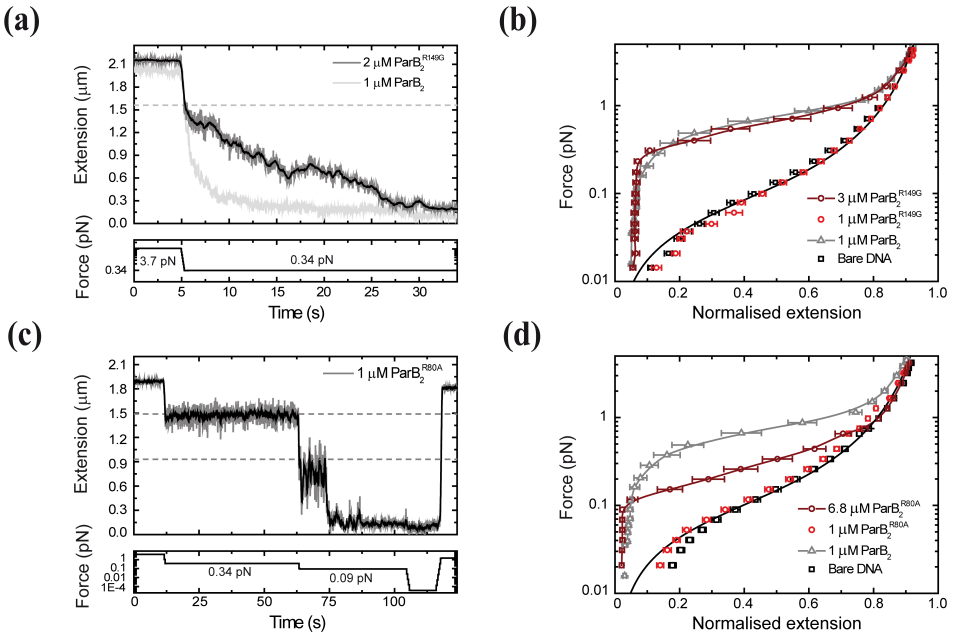

Figure 5.2: ParB condensation is independent of specific *parS* binding and dependent on the N-terminal domain. a) Representative condensation traces for wild type ParB and ParB^{R149G}. ParB was introduced at 3.7 pN. Expected DNA extension in the absence of ParB indicated by dashed lines. b) Mean force-extension curves (FECs) for *parS*-containing DNA molecules in the presence ParB^{R149G}. c) Typical condensation trace of ParB^{R80A} showing condensation events at weak stretching forces. d) Mean FECs for *parS*-containing DNA molecules in the presence ParB^{R80A}. Condensation Non-condensed DNA data is fitted to the worm-like chain model (WLC). Solid lines in condensed data are guides for the eye. Errors are the standard error of the mean (s.e.m.) of measurements on different molecules (N∼15-35 molecules).

overlapping FECs. These results are in agreement with our previous results, which showed similar condensation efficiencies with ParB on *parS* and non-specific DNA substrates [Taylor et al., 2015]. This is consistent with the idea that nsDNA binding may occur at a second DNA binding locus.

To determine the role of the NTD in DNA compactation we tested the activity of the spreading-defficient mutant R80A. This mutation has been described to inhibit the formation of foci *in vivo*, while both *parS* and non-specific binding is retained [Autret et al., 2001]. MT condensation assays show that ParB^{R80A} does not condense DNA molecules at the typical conditions the wild-type protein condenses the DNA (Figure 5.2c). However, at highly reduced forces there are condensation events and these are usually abrupt and coupled to a substantial drop of extension. Increasing the concentration of ParB$_2^{R80A}$ up to 6.8 μM allows condensation at higher forces, but condensation was highly depleted compared to wild-type condensation, as shown by the lower condensation forces reported in FECs (Figure 5.2d). Since the R80A mutation does not significantly affect DNA binding [Autret et al., 2001; Song et al., 2017], MT experiments allude to an implication of the NTD in the formation of protein-protein interfaces.

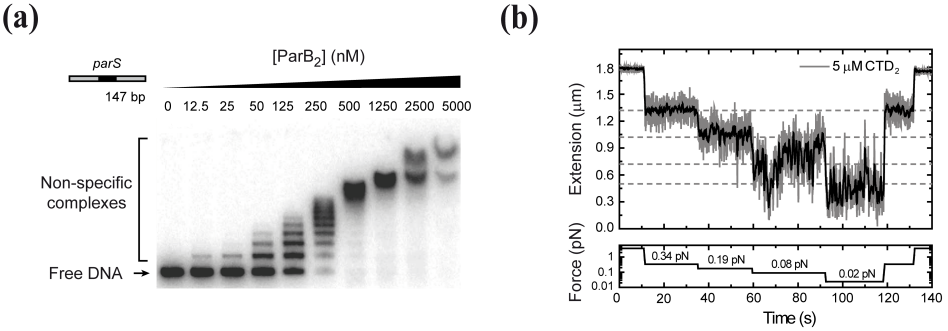

Figure 5.3: The CTD of ParB binds but does not condense DNA. EMSA assay for the titration of full length ParB and CTD against 147 bp DNA. EMSA assays performed by Gemma L.M. Fisher. b) Magnetic tweezers traces of extension as a result of the indicated stretching forces. No condensation events are observed. The expected extension in absence of ParB at each force is indicated by dashed lines.

5.3.2 The CTD binds non-specifically via lysine-rich surface

Our previous results led us to test the idea that the CTD might be responsible of binding to non-specific DNA. We performed EMSA assays with purified CTD. These showed that the CTD was indeed able to bind DNA (Figure 5.3a) resulting in the formation of a ladder of bands of decreasing mobility. MT experiments indicate that, unlike full length ParB, the CTD was not capable of condensing DNA tethers under any condition tested, even up to 5 μM CTD_2 concentrations and under applied forces as low as 0.02 pN (Figure 5.3b). This is consistent with the expected requirement for multiple protein-protein and/or protein-DNA interfaces to promote DNA looping and condensation.

To identify the residues potentially implicated in DNA binding, we constructed a homology modelling structure of the *B. subtilis* CTD. The modelled structure was aligned with the P1 ParB structure (PDB entry: 2NTZ) from *E. coli* which was described to interact with nsDNA by the CTD [Schumacher et al., 2007]. A lysine-rich region was found between residues 252 and 257, and aminoacids at positions 255 and 257 were identified as the main candidates for non-specific binding to DNA.

To determine the role of the lysine residues in DNA binding we substituted lysine residues at positions 255 and 257 to alanines. We monitored the effect on DNA binding of the the dual substitution K255A/K257a in the context of both the full length ParB protein ($ParB^{KK}$) and the CTD-only truncation (CTD^{KK}). EMSA assays with the full-length ($ParB^{KK}$) variant showed no deffect in specific binding to *parS* [Fisher et al., 2017]. However, the double lysine mutant showed similar DNA binding affinity than wild-type ParB to non-specific DNA [Fisher et al., 2017]. CTD^{KK} displayed a greatly reduced affinity ($\sim$ 50-fold) for DNA, but the binding was not completely abolished [Fisher et al., 2017].

We next designed a triple K252A/K255A/K259A variant with the aim of fully dissipating the positive charge density across the surface of the CTD. Interestingly, EMSA analysis of full length $ParB^{KKK}$ showed a clear and consistent defect in nsDNA bind-

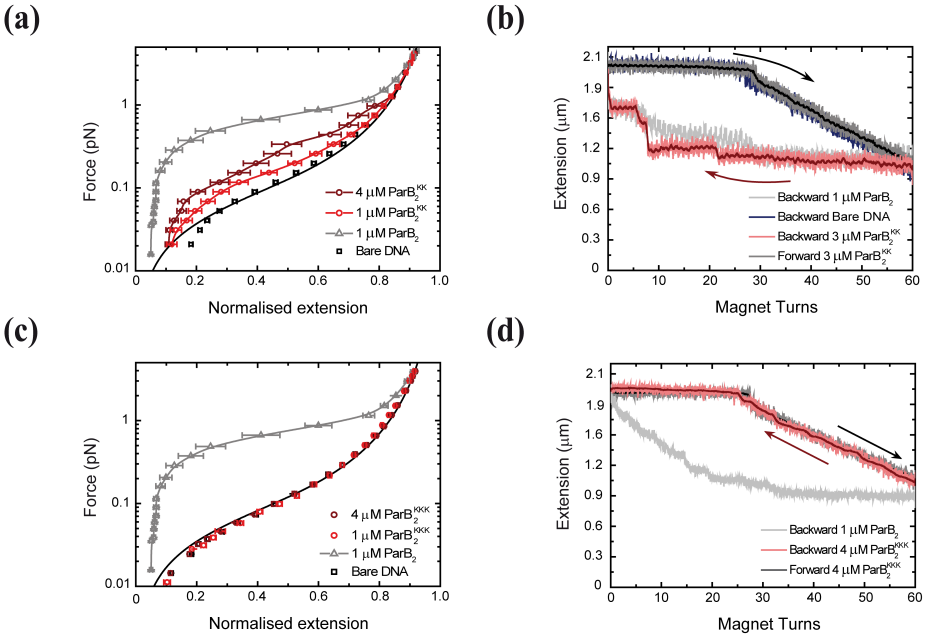

Figure 5.4: DNA binding by the CTD is required for efficient DNA condensation. a) Mean force-extension curves (FEC) of DNA molecules co-incubated with $ParB^{KK}$. Non-condensed (protein-free) DNA data is fitted to the worm-like chain model. Solid lines in condensed data are guides for the eye. Errors are the standard error of the mean of measurements on different molecules (N∼18-35 molecules). b) Plectoneme stabilisation assay comparing bare DNA, wild-type ParB and $ParB^{KK}$. The double-mutant protein supported DNA bridging and occasionally large steps were observed in the backward trace. c) FECs of DNA molecules co-incubated with $ParB^{KKK}$. d) Plectoneme stabilisation assay comparing wild type ParB and $ParB^{KKK}$. No activity was detected for the triple-mutant protein.

ing while the ParB-*parS* complexes were still formed [Fisher et al., 2017]. We argue that the retained non-specific binding activity found could reflect a weak ability of the CTBD to bind nsDNA independently of the CTD. This idea is consistent with the full inhibition of nsDNA binding found in the truncated variant CTD^{KKK} [Fisher et al., 2017].

5.3.3 DNA binding by the CTD is essential for DNA condensation and bridging

The double- and triple-lysine mutant ParB proteins were tested for their ability in forming condensed ParB networks. Incubation of full length 1 μM $ParB_2^{KK}$ with single DNA tethers resulted in defective DNA condensation compared to wild type ParB as reported by force-extension data (Figure 5.4a). When it was observed, condensation was sudden and it required to drop the applied force to exceptionally low values (< 0.3 pN), similarly as occurred for $ParB_2^{R80A}$. The DNA molecules also showed unusually large steps when decondensed by force, suggesting that $ParB^{KK}$ was infrequently stabilising *in trans* DNA-bridging interactions between isolated DNA regions (data

not shown). At higher concentrations, 4 μM $ParB_2^{KK}$, condensation was still highly inefficient and when present in the form of the previously described abrupt reduction of extension. The frequency and the extension of the condensastion events increased at higher concentrations, and started at forces up to 1 pN. Nonetheless, condensation was still clearly depleted when compared with wild-type condensation and in every case not present in all the DNA molecules.

We also performed plectoneme stabilisation experiments as described in section 4.3.3 (Figure 4.12). In this assay, a single torsionally-constrained DNA molecule is positively supercoiled at 4 pN force by applying 60 turns. ParB is then introduced and, after full buffer exchange, all turns are released whilst monitoring DNA extension. Any reduction of DNA extension from that expected for bare DNA is indicative of supercoiled regions being stabilised by ParB. For concentrations above 3 μM, we found that $ParB^{KK}$ could strongly stabilise DNA-bridging interactions (Figure 5.4b). However, the release of bridging during plectoneme removal is characterised by large steps in extension, indicating a low number of proteins involved in the interaction, in agreement with the results obtained from EMSA assays.

Remarkably, co-incubation of $ParB^{KKK}$ with single DNA tethers under our standard experimental conditions resulted in no measurable condensation events, at elevated concentrations and low forces (Figure 5.4c). In addition, plectoneme stabilisation assays does not show any sign *in cis* of interaction, as reported by the overlap of the backward trace with that obtained in bare DNA (Figure 5.4d).

5.3.4 The CTD can both inhibit the formation and decondense ParB-DNA networks

The CTD potentially acts as both a dimerisation interface and as site for nsDNA-binding. Therefore, we hypothesised that the CTD might have a dominant negative effect on full length ParB by competing for the DNA and protein interfaces that mediate the formation of ParB networks in MT.

Purified CTD completely inhibited the formation of condensed networks if pre-incubated in a five-times excess with wild-type ParB (Figure 5.5a). To test the idea that the CTD dimerisation interface is also important for maintaining the condensed state, we repeated our experiments with the CTD^{KK} and CTD^{KKK} constructs. Co-incubation of wild-type ParB with the non-specific DNA binding mutant CTD^{KKK} inhibited DNA condensation by full length wild-type with a similar efficiency to that of wild-type CTD (Figure 5.5b). Similarly, the CTD^{KK} truncation –which we reported as highly deffective in nsDNA binding– was also efficient inhibiting DNA binding (Figure 5.5c).

We also tested whether the introduction of free CTD to pre-condensed tethers was able to disrupt ParB-DNA networks. First, we condense DNA molecules incubated with 1 μM $ParB_2$ by reduction of the force from 4 pN to 0.34 pN. After condensation is triggered, 1μM $ParB_2$ co-incubated with 5 μ CTD_2 is injected in the flow cell (Figure 5.6a). The inclusion of an excess of free CTD rapidly disrupted ParB networks recovering the expected extension for 0.3 pN in absence of protein. Some degree

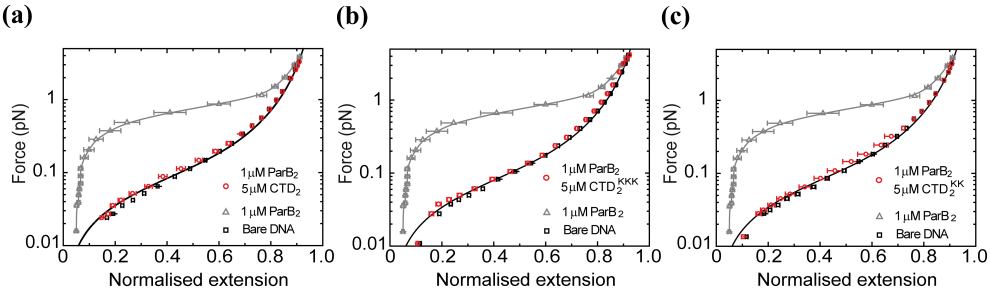

Figure 5.5: The CTD of ParB inhibits ParB-dependent condensation. a) Mean force-extension curves (FEC) for DNA molecules co-incubated with 1 μM $ParB_2$ in the presence or absence of 5 μM CTD_2. b) Mean FECs for DNA molecules co-incubated with 1 μM $ParB_2$ in the presence or absence of 5 μM mutant CTD_2^{KKK}. c) Mean FECs for DNA molecules co-incubated with 1 μM $ParB_2$ in the presence or absence of 5 μM mutant CTD_2^{KK}. Errors are s.e.m. of measurements on different molecules (N>40 molecules). Bare DNA data is fitted to the worm-like chain model. Solid lines in condensed data are guides for the eye.

of decondensation was observed in 94% of all the molecules tested. Incubation in absence of force for ∼30 seconds did not promote stabilisation of ParB networks, as indicated by the absence of decondensation patterns after increasing the force to 4 pN. The release of condensed ParB networks also occurs when 1μM $ParB_2$ and CTD^{KKK} are introduced together in the flow cell (Figure 5.6b). Similar results were obtained with the mutant CTD^{KK}. Control experiments injecting $ParB_2$ in absence of CTD truncations shows that decondensation is not induced by the applied drag force of the flow –drag force was estimated to be $\sim$ 0.02 pN, see Appendix C– and that the condensed state is stable at the time scale of the experiments (Figure 5.6c). Furthermore, DNA tethers were also able to re-condense following force-induced decondensation, in large contrast with the experiments in the presence of the CTD-only truncation.

The ability of the CTD to decondense ParB networks demonstrates that the protein-protein and/or protein:DNA interfaces that maintain the condensed state under a low force regime are dynamic i.e. they are exchanging while the overall structure of the network is maintained. Analysis of condensation ratios (see materials and methods) shows a decondensation efficiency $>$ 90% for all the CTD variants (Figure 5.6d). Remarkably, mutations impairing DNA binding by the CTD do not affect the inhibitory activity of the CTD in ParB wild-type condensation. Thus, these results indicate that CTD competition is not at the level of DNA binding, but at the level of protein-protein interactions. Consequently, this result evidences a dependency on the formation of ParB-ParB interactions mediated by the CTD for efficient DNA condensation.

5.4 Discussion

ParB proteins form long-distance bridging interactions on DNA substrates to form foci that facilitate chromosomal partitioning reactions [Bingle et al., 2005; Lynch and Wang, 1995; Rodionov et al., 1999]. These ParB foci are anchored at *parS* sites and in-

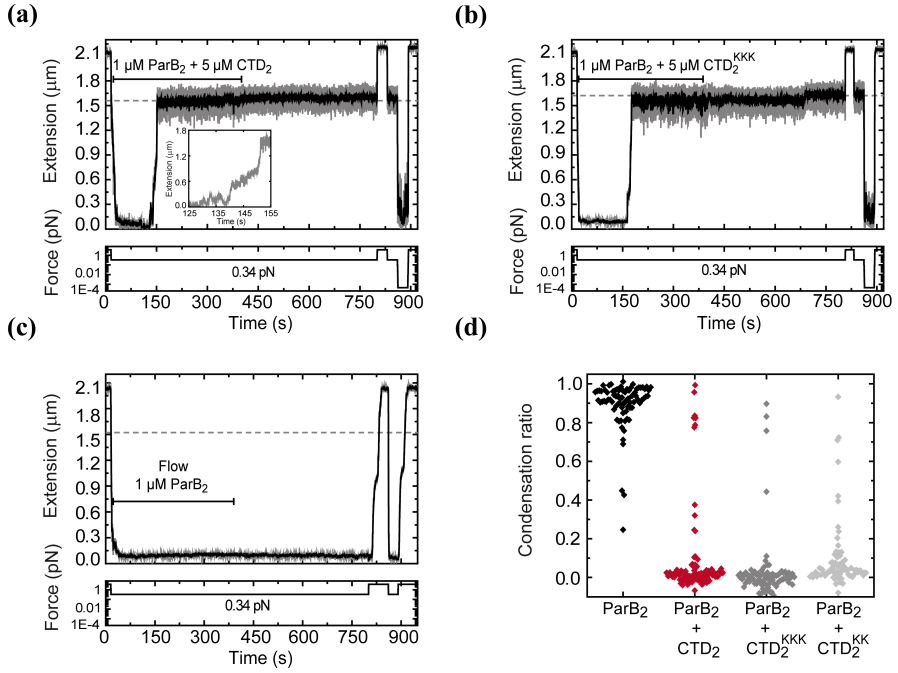

Figure 5.6: The CTD of ParB disrupts ParB condensed networks. a) ParB-DNA complexes decondense following the introduction of 1 μM $ParB_2$ and 5 μM free CTD_2. b) ParB-DNA complexes decondense after injection of 1 μM $ParB_2$ and 5 μM CTD_2^{KKK}. c) DNA molecules condensed by ParB are not decondensed by drag and are stable in magnetic tweezers in the presence of 1 μM $ParB_2$. d) Condensation ratio for individual DNA condensation events involving the addition of CTD competitor variants to pre-condensed ParB-DNA networks. $N > 40$ molecules per variant.

teract non-specifically around a single site for ± 18 kbp [Breier and Grossman, 2007; Wang et al., 2017]. This ParB *spreading* activity appears to be a conserved property across chromosomal and plasmid segrosomes, yet the protein interaction interfaces involved have remained elusive, particularly for the genomically-encoded systems. This is, in part, due to the highly variable structures of ParB proteins and their cognate centromere sequences, even within the type I subclass of which *B. subtilis* ParB is a member [Gerdes et al., 2000; Schumacher, 2008]. Increasing evidence indicates that ParB spreading is the result of a DNA-bridging activity mediated by ParB-ParB oligomerisation interfaces [Broedersz et al., 2014; Graham et al., 2014; Sanchez et al., 2013, 2015; Taylor et al., 2015]. For the type I ParB protein class, there is evidence to suggest that dimerisation and/or tetramerisation can occur at the N-terminal domain (NTD) and central DNA-binding domain (CDBD), and that dimerisation can occur at the C-terminal domain (CTD) [Chen et al., 2015; Leonard et al., 2004; Schumacher and Funnell, 2005]. A combination of some or all of these activities must support ParB oligomerisation.

We find that the CTD of ParB binds non-specific DNA. A positively charged Lys/Arg-rich surface located at the CTD of P1 ParB has been shown to bind directly

to DNA using structural and/or biochemical techniques [Schumacher and Funnell, 2005; Schumacher et al., 2007]. To directly address the putative role of the CTD in spreading [Leonard et al., 2004], we rationally designed mutations located at the positively charged region of the CTD. These mutants allowed us to separately test the role of dimerisation and DNA binding domains in ParB function. We showed that the CTD is not involved in *parS* recognition, and that this is instead dependent on the HtH motif found within the CDBD domain as predicted in several previous studies [Autret et al., 2001; Bignell and Thomas, 2001; Chen et al., 2015; Gerdes et al., 2000; Graham et al., 2014; Gruber and Errington, 2009; Lobocka and Yarmolinsky, 1996; Theophilus and Thomas, 1987]. In contrast, the CTD is essential for the formation of nsDNA complexes. Mutants at the lysine-rich patch were defective in non-specific DNA binding and condensation. Foremost, we find an inhibition of wild-type condensation and a disruption of ParB networks when co-incubated with an excess of free CTD. The inhibitory activity of ParB condensation by free CTD is independent on the ability of the CTD to bind DNA. These observations strongly support the idea that the CTD is essential for both DNA binding and for establising ParB-ParB bridging interactions for condensation.

Strikingly, our results reveal an additional contribution of the NTD for DNA condensation. A single mutation located at the Box II of the NTD (R80A) strongly affected DNA condensation. This mutation has been described as not interfering with nsDNA binding [Autret et al., 2001; Song et al., 2017]. Hence, the weak condensation reported for this mutant cannot be attributed to a defective nsDNA binding, but to an impaired spreading activity due to limited protein-protein interactions. This result is in agreement with recent structural and single-molecule studies of *H. pylori* and *B. subtilis* ParB, evidencing the importance in bridging of a novel tetramerisation interface within the NTD [Chen et al., 2015; Song et al., 2017].

The average work done by ParB and variants during condensation was determined from the difference between the integral of the force-extension curve in the presence of the protein and that of DNA alone. This provides a means to quantitatively compare the condensation efficiency between mutants and summarises the results presented in this chapter (Figure 5.7). Together, these results illustrate the complexity of ParB activity, where DNA condensation is dependent on the coordinated interplay between the NTD-CDBD and the CTD for both DNA binding, spreading and bridging. In addition, these data set the basis for understanding the energetics of ParB condensation. At the typical condensation force (0.34 pN), the vertical resolution in our MT setup is $\sigma_z = 80$ nm, and thus we should be able to easily resolve condensation events of $\Delta z = 2\sigma_z$ nm. Conversely, MT condensation experiments evidence an absence of well-defined steps during ParB wild-type condensation, indicating that actually $\Delta z < 2\sigma_z$. Full condensation of the DNA tethers involves an extension reduction of $\Delta l = 2$ µm, involving the interaction of at least $\langle N \rangle = \Delta l / 2\sigma_z = 12$ protomers. We obtained an average work during ParB condensation of $\Delta W = 930$ pN nm, setting an upper limit for the energy between ParB protomers to $\epsilon_{\text{ParB}} = \Delta W / \langle N \rangle = 77$ pN nm ($\sim 20 k_B T$). We anticipate that this result will be a reference parameter for simulation and modelling of ParB condensation.

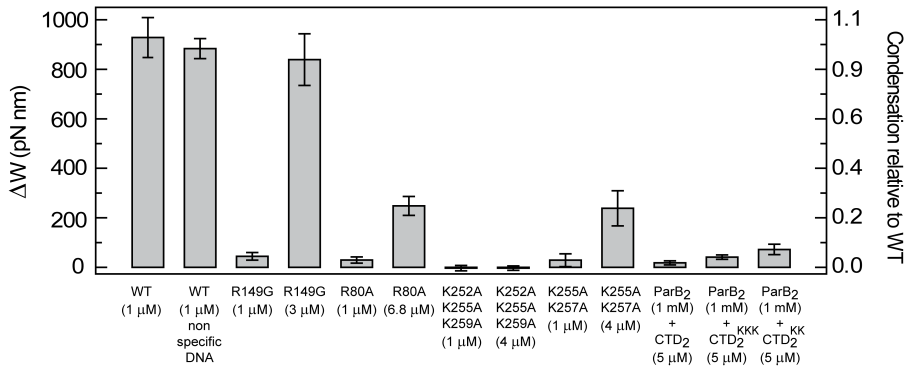

Figure 5.7: Condensation work (ΔW) by ParB variants. Mutations at both NTD and CTD strongly impair DNA condensation. Co-incubation of CTD variants with ParB prevents the formation of condensed networks. Error bars are s.e.m. for $N > 10$ DNA molecules.

The presence of two independent DNA binding and dimerisation loci in ParB can help to explain how ParB networks are anchored at *parS in vivo*. In cells, the pool of ParB has been estimated in ∼440 ParB dimers per cell [Graham et al., 2014] or 5.8 μM $ParB_2$ (from a typical *B. subtilis* cell volume of 4.6 μm^3, [Yu et al., 2013]). This concentration is on same range to those employed in this work. Furthermore, the length of the *B. subtilis* chromosome (4.2 Mbp) is within the order of magnitude of the expected number of non-specific binding sites in our magnetic tweezers experiments, since an individual DNA tether already contributes to 2 kbp. Moreover, the bacterial chromosome is highly occupied by DNA processing machines, and thus the number of available sites for ParB binding is reduced. This reasoning allow us to speculate that the conditions in *in vivo* are suitable for non-specific binding and condensation. This presents a conundrum to explain the strict localization of ParB around *parS* sites from the apparently low specificity for *parS in vitro* ($<$ 10-fold greater affinity for *parS* versus nsDNA) [Taylor et al., 2015]. It has been proposed that spreading could be facilitated by *parS*-dependent conformational changes that act as nucleation points for networks [Broedersz et al., 2014; Leonard et al., 2004]. However, this hypothesis still does not explain why *parS* is not required for DNA condensation *in vitro* [Graham et al., 2014; Taylor et al., 2015], whereas the absence of *parS* sites prevents the formation of ParB-DNA foci *in vivo* [Erdmann et al., 1999; Graham et al., 2014; Sanchez et al., 2015]. Monte-Carlo simulations suggested that – for a ParB system composed of two dimerisation domains and a single DNA binding domain– a mix of spreading and bridging is the minimal model for the formation of localized condensed complexes [Broedersz et al., 2014]. The presence of separated DNA binding domains for specific and non-specific binding might aid the formation of complexes around *parS* sequences, even with the apparent low specific affinity for the *parS* sequence.

Nevertheless, additional factors could play a determinant role in ParB specificity for *parS in vivo*, as has been shown for homologous ParB systems. For instance, the P1 ParB (*E. coli*) shows a strong enhancement of *parS* recognition after protein-

induced DNA bending by IHF [Funnell, 1991; Funnell and Gagnier, 1993]. Alternatively, nucleoid associated proteins might constrain the number of available non-specific binding sites for ParB, favoring the binding around *parS* sites. A more complete understanding of ParB specific binding and network formation will be required to underpin future studies on how ParB acts together with ParA and condensin to orchestrate efficient chromosome segregation.

References

Autret, S., Nair, R., and Errington, J. Genetic analysis of the chromosome segregation protein Spo0J of bacillus subtilis: evidence for separate domains involved in DNA binding and interactions with Soj protein. *Mol. Microbiol.*, 41(3):743–55, 2001.

Bignell, C. and Thomas, C. M. The bacterial ParA-ParB partitioning proteins. *J. Biotechnol.*, 91(1):1–34, 2001.

Bingle, L. E., Macartney, D. P., Fantozzi, A., Manzoor, S. E., and Thomas, C. M. Flexibility in repression and cooperativity by KorB of broad host range IncP-1 plasmid RK2. *J. Mol. Biol.*, 349(2):302–16, 2005.

Bouet, J. Y. and Funnell, B. E. P1 ParA interacts with the P1 partition complex at parS and an ATP-ADP switch controls parA activities. *EMBO J.*, 18(5): 1415–24, 1999.

Breier, A. M. and Grossman, A. D. Whole-genome analysis of the chromosome partitioning and sporulation protein Spo0J-(ParB) reveals spreading and origin-distal sites on the bacillus subtilis chromosome. *Mol. Microbiol.*, 64(3): 703–18, 2007.

Broedersz, C. P., Wang, X., Meir, Y., Loparo, J. J., Rudner, D. Z., and Wingreen, N. S. Condensation and localization of the partitioning protein ParB on the bacterial chromosome. *Proc. Natl. Acad. Sci. U.S.A.*, 111(24): 8809–14, 2014.

Chen, B. W., Lin, M. H., Chu, C. H., Hsu, C. E., and Sun, Y. J. Insights into ParB spreading from the complex structure of Spo0J and parS. *Proc. Natl. Acad. Sci. U.S.A.*, 112(21):6613–8, 2015.

Davey, M. J. and Funnell, B. E. Modulation of the P1 plasmid partition protein ParA by ATP, ADP, and P1 ParB. *J. Biol. Chem.*, 272(24):15286–92, 1997.

Davis, M. A., Martin, K. A., and Austin, S. J. Biochemical activities of the para partition protein of the P1 plasmid. *Mol. Microbiol.*, 6(9):1141–7, 1992.

Erdmann, N., Petroff, T., and Funnell, B. E. Intracellular localization of P1 ParB protein depends on ParA and parS. *Proc. Natl. Acad. Sci. U.S.A.*, 96(26): 14905–10, 1999.

Fisher, G. L. M., Pastrana, C. L., Higman, V. A., Koh, A., Taylor, J. A., Butterer, A., Craggs, T. D., Sobott, F., Murray, H., Crump, M. P., Moreno-Herrero, F., and Dillingham, M. S. The C-terminal domain of ParB is critical for dynamic DNA binding and bridging interactions which condense the bacterial centromere. *bioRxiv*, 2017.

Funnell, B. E. The P1 plasmid partition complex at parS. the influence of Escherichia coli integration host factor and of substrate topology. *J. Biol. Chem*, 266(22):14328–37, 1991.

Funnell, B. E. and Gagnier, L. The P1 plasmid partition complex at parS. II. Analysis of ParB protein binding activity and specificity. *J. Biol. Chem*, 268(5): 3616–24, 1993.

Gerdes, K., Møller-Jensen, J., and Bugge Jensen, R. Review plasmid and chromosome partitioning: surprises from phylogeny. *Mol. Microbiol.*, 37(3):455–466, 2000.

Graham, T. G., Wang, X., Song, D., Etson, C. M., van Oijen, A. M., Rudner, D. Z., and Loparo, J. J. ParB spreading requires DNA bridging. *Genes Dev.*, 28 (11):1228–38, 2014.

Gruber, S. and Errington, J. Recruitment of condensin to replication origin regions by ParB/Spo0J promotes chromosome segregation in B. subtilis. *Cell*, 137(4):685–696, 2009.

Leonard, T., Butler, P., and Löwe, J. Structural analysis of the chromosome segregation protein Spo0J from Thermus thermophilus. *Mol. Microbiol.*, 53 (2):419–432, 2004.

Lin, D. C. and Grossman, A. D. Identification and characterization of a bacterial chromosome partitioning site. *Cell*, 92(5):675–685, 1998.

Lobocka, M. and Yarmolinsky, M. P1 plasmid partition: a mutational analysis of ParB. *J. Mol. Biol.*, 259(3):366–82, 1996.

Lynch, A. S. and Wang, J. C. SopB protein-mediated silencing of genes linked to the sopC locus of Escherichia coli F plasmid. *Proc. Natl. Acad. Sci. U.S.A.*, 92(6):1896–900, 1995.

Murray, H., Ferreira, H., and Errington, J. The bacterial chromosome segregation protein Spo0J spreads along DNA from parS nucleation sites. *Mol. Microbiol.*, 61(5):1352–61, 2006.

Radnedge, L., Youngren, B., Davis, M., and Austin, S. Probing the structure of complex macromolecular interactions by homolog specificity scanning: the P1 and P7 plasmid partition systems. *EMBO J.*, 17(20):6076–85, 1998.

Rodionov, O., Lobocka, M., and Yarmolinsky, M. Silencing of genes flanking the P1 plasmid centromere. *Science*, 283(5401):546–9, 1999.

Sanchez, A., Rech, J., Gasc, C., and Bouet, J. Y. Insight into centromere-binding properties of ParB proteins: a secondary binding motif is essential for bacterial genome maintenance. *Nucleic Acids Res.*, 41(5):3094–103, 2013.

Sanchez, A., Cattoni, D. I., Walter, J. C., Rech, J., Parmeggiani, A., Nollmann, M., and Bouet, J. Y. Stochastic self-assembly of ParB proteins builds the bacterial DNA segregation apparatus. *Cell Syst.*, 1(2):163–73, 2015.

Schumacher, M. A. Structural biology of plasmid partition: uncovering the molecular mechanisms of DNA segregation. *Biochem. J.*, 412(1):1–18, 2008.

Schumacher, M. A. and Funnell, B. E. Structures of ParB bound to DNA reveal

mechanism of partition complex formation. *Nature*, 438(7067):516–9, 2005.

Schumacher, M. A., Mansoor, A., and Funnell, B. E. Structure of a four-way bridged ParB-DNA complex provides insight into P1 segrosome assembly. *J Biol Chem.*, 282(14):10456–64, 2007.

Song, D., Rodrigues, K., Graham, T. G., and Loparo, J. J. A network of cis and trans interactions is required for ParB spreading. *Nucleic Acids Res.*, 45(12): 7106–7117, 2017.

Taylor, J. A., Pastrana, C. L., Butterer, A., Pernstich, C., Gwynn, E. J., Sobott, F., Moreno-Herrero, F., and Dillingham, M. S. Specific and non-specific interactions of ParB with DNA: implications for chromosome segregation. *Nucleic Acids Res.*, 43(2):719–731, 2015.

Theophilus, B. D. and Thomas, C. M. Nucleotide sequence of the transcriptional repressor gene korb which plays a key role in regulation of the copy number of broad host range plasmid RK2. *Nucleic Acids Res.*, 15(18):7443–50, 1987.

Wang, X., Tang, O. W., Riley, E. P., and Rudner, D. Z. The SMC condensin complex is required for origin segregation in Bacillus subtilis. *Curr. Biol.*, 24(3):287–92, 2014.

Wang, X., Brandão, H. B., Le, T. B., Laub, M. T., and Rudner, D. Z. Bacillus subtilis SMC complexes juxtapose chromosome arms as they travel from origin to terminus. *Science*, 355(6324):524–7, 2017.

Yu, A. C., Loo, J. F. C., Yu, Samuel Kong, S. K., and Chan, T.-F. Monitoring bacterial growth using tunable resistive pulse sensing with a pore-based technique. *Appl. Microbiol. Biotechnol.*, 98(2):855–862, 2013.

Part III

Initiation of the Rolling-circle replication: example of a topology-dependent reaction

6

Force and twist dependence of RepC nicking activity

Chapter overview. Many bacterial plasmids replicate by an asymmetric rolling-circle mechanism that requires sequence-specific recognition for initiation, nicking of one of the plasmid DNA strands, and unwinding of the duplex prior to subsequent DNA synthesis. Nicking is performed by a replication-initiation protein (Rep) that directly binds to the plasmid double-stranded origin and remains covalently bound to its 5'-end via a phosphotyrosine linkage. It has been proposed that the inverted DNA sequences at the nick site form a cruciform structure that facilitates DNA cleavage. However, the role of Rep proteins in the formation of this cruciform and the implication for its nicking and religation functions is unclear. In this chapter, the DNA nicking and religation activities of RepC are presented as well as their implication in the initiation of replication.

This chapter has been published by Cesar L. Pastrana, Carolina Carrasco,..., and Fernando Moreno-Herrero in *Nucleic Acids Research*, 2016, 44(18): 8885-96.

6.1 Introduction

6.1.1 Rolling-Circle replication of plasmids

Plasmids play a crucial role in prokaryotic evolution by transferring DNA between bacterial species. The incorporation of DNA plasmids in the host cell commonly involves a gain of function for the recipient organism. For instance, resistance to antibiotics and heavy metals are encoded in plasmids, and thus they have important consequences for ecology and human health. There are three main replication mechanisms for circular plasmids: theta-type, strand displacement and rolling-circle [del Solar et al., 1998]. Rolling-Circle Replication (RCR) is the most common mode of replication for small ($<$10 kbp) plasmids in Gram-positive bacteria [Khan, 1997a; Ruiz-Masó et al., 2015]. RCR is initiated by replication initiator proteins (Rep proteins) encoded by the plasmid. The origins of replication of the plasmid contain the binding sites of the Rep proteins as well as their nicking sites [Khan, 1996; Koepsel et al., 1985a]. Rep proteins are able to nick and ligate DNA through topoisomerase type I mechanism [Ruiz-Masó et al., 2015; Khan, 1997b].

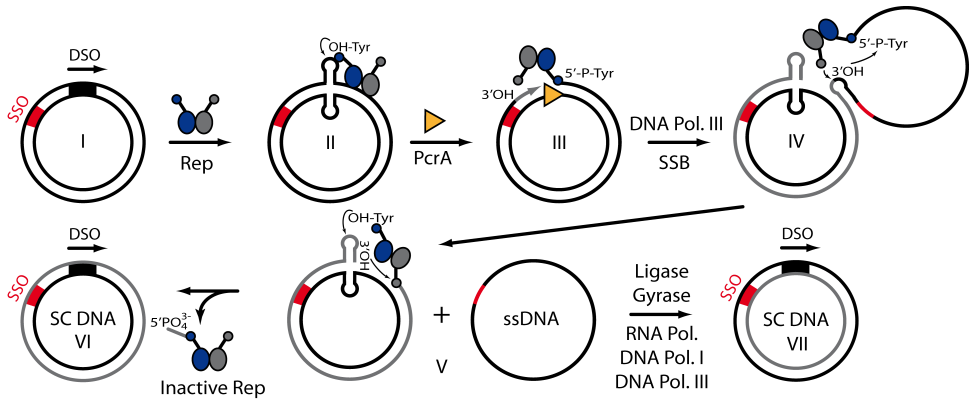

Figure 6.1: Model of plasmid rolling-circle replication. Step I) A supercoiled plasmid containing the double-strand origin (dso) and single-strand origin (*sso*) is the substrate for DNA replication and is first nicked by a Rep dimer. Step II) A secondary structure at the *dso* facilitates binding and/or nicking of the DNA by RepC which becomes covalently attached to the 5'-PO_4 end of the DNA after nicking. Step III) PcrA helicase interacts with RepC to initiate the unwinding of the duplex DNA. Step IV) DNA polymerase III loads at the free 3'OH generated after nicking and synthesizes a new leading strand, generating a ssDNA plasmid after full replication. Then, the not bound RepC monomer nicks at the regenerated *dso*. Step V) This activity generates a nicked dsDNA plasmid with RepC covalently bound and a single-stranded DNA (ssDNA) plasmid. Step IV) The nicked plasmid is sealed from the 3'OH attack of the DNA to the phosphate of the RepC-DNA bond. Step VII) The covalently closed ssDNA is converted to dsDNA by the synthesis of an RNA primer at the *sso* by the RNA polymerase followed by DNA synthesis by DNA polymerases I and III.

RepC is a 38-kDa homodimeric protein encoded by the pT181 plasmid of *Staphylococus aureous*. It binds to a specific DNA sequence contained within the double-stranded origin of replication (*dso*) of the plasmid (Figure 6.1, step I) and creates a nick (Figure 6.1, step II) [Koepsel et al., 1985b; Jin et al., 1997]. The nicking site is contained within an inverted repeat that is likely to form a cruciform structure [Noirot et al., 1990]. RepC nicks the leading strand and becomes covalently bound to its 5'-end via a phosphotyrosine linkage. Then, PcrA helicase loads onto the short length of accessible single-stranded DNA (ssDNA) [Noirot et al., 1990; Arbore et al., 2012] (Figure 6.1, step III). The free 3'-end at the nick is used for extension synthesis where PcrA, DNA polymerase III, and single-strand DNA binding protein (SSB) coordinate their activities at the origin to replicate the entire plasmid [Thomas et al., 1990]. Replication of DNA by DNA polymerase III proceeds until the plasmid is fully unwound and a new leading strand has been fully synthesized. Then, the not-bound monomer of RepC cleaves the displaced ssDNA at the regenerated nick site at the junction of the old and newly synthesized leading strands (Figure 6.1, step IV)[Chang et al., 2000]. The resulting 3'OH on the DNA after RepC nicking acts as nucleophile, attacking the phosphate involved in the DNA-RepC bond. This generates a nicked plasmid of original and new synthesis strands, as well as a closed ssDNA corresponding to the complementary original strand (Figure 6.1, step V). At the termination step,

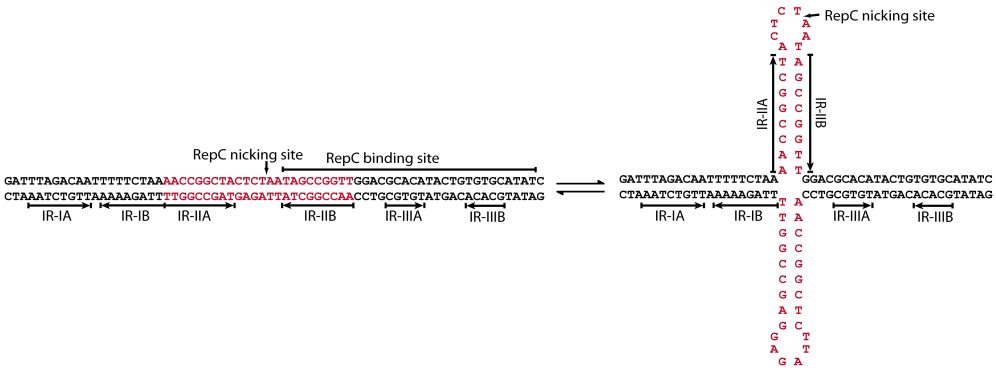

Figure 6.2: Double strand origin (*dso*) of replication of the pT181 plasmid. The origin region contains the RepC binding and nicking sites and three sets of inverted repeats (IR). The red area comprising the IR-II is thought to form a cruciform structure, positioning the RepC nicking site at the tip.

in a process that is not fully understood, the RepC monomer not-bound to the DNA nicks at the tip of the cruciform. This generates a free 3'OH group, which reacts with the phospotyrosine bond of one of the RepC monomers bound to DNA following a similar mechanism to that described for the step IV, and thus the nick in the plasmid is sealed. Religation results in RepC dissociation, leaving a relaxed, closed circular double-stranded DNA (dsDNA). The relaxed DNA is supercoiled (SC) by DNA gyrase while the RepC dimer is released from the DNA in an inactive form, covalently bound to a short ssDNA fragment from one of the cruciform arms (Figure 6.1, step VI)[Rasooly and Novick, 1993]. The displaced single leading strand is then converted to the double-stranded form by using the single strand origin of replication (*sso*), and requires of the RNA polymerase I for primer synthesis and DNA polymerases I and III for replication (Figure 6.1, step VII) [Khan, 1997a; Gruss et al., 1987; Kramer et al., 1985].

Formation of a cruciform structure at the *dso* has long been proposed, as it contains multiple inverted repeats at this region (Figure 6.2) [Gennaro et al., 1989; Noirot et al., 1990]. However, the role of this potential secondary structure of DNA for RepC and PcrA binding and/or activity is still unclear. Rep proteins have been proposed to induce the formation of a cruciform structure at the *dso* binding site [Jin et al., 1997; Noirot et al., 1990; Koepsel et al., 1986b]. This model of cruciform formation could be considered as active in that the binding of the protein melts the DNA and actively extrudes the cruciform structure. We instead propose here that RepC plays a passive role in this process and that cleavage only occurs when particular conditions of force and twist on the DNA are fulfilled.

To quantitatively investigate the roles of DNA force and twist in RCR, we have used Magnetic Tweezers (MT) to directly measure the nicking activity of RepC in several kilo basepair long DNA molecules with different degrees of supercoiling and at different applied forces. RepC exclusively nicked negatively supercoiled DNAs and this activity was force-dependent. We have also used MT to measure the relaxation velocity of supercoiled DNA upon nicking and compared RepC data with that

obtained for the human TopIB (hTopIB) and the nicking enzyme Nb.BbvCI. hTopIB and Nb.BbvCI represent two extreme cases of strong and mild interaction with DNA during plectoneme release as the former fully encircles the DNA duplex [Redinbo et al., 1998] and the latter is expected to release from its substrate after nicking. Plectonemes were released by RepC at a velocity intermediate of that produced by hTopIB and Nb.BbvCI, suggesting a different mode on interaction compared to the other two enzymes. The frequency of religation was also compared between hTopIB and RepC. RepC behaved as an inefficient religating enzyme highlighting its role as a replication initiator instead of as a topoisomerase.

Based on the work described here and structural predictions, we propose a model where RepC nicking activity is force- and twist-dependent. Nicking of DNA at the *dso* by RepC is passive and a consequence of binding to a substrate with the appropriate degree of torque, in which the cruciform structure required for nicking is already present, rather than promoted as a result of RepC binding.

6.2 Materials and methods

6.2.1 Proteins

The MBP-RepC fusion protein, in which the maltose binding protein (MBP) is fused to the N- terminal end of RepC, was purified[1] as described earlier [Chang et al., 2000].

6.2.2 DNA substrates

The pBlueScriptIISK:pT181*cop608* (7217 bp) plasmid was generated by ligating the pBlueScriptIISK+ plasmid (Stratagene) and the pT181*cop608* plasmid (4257 bp) at their KpnI sites [Khan and Novick, 1983]. DNA constructs for magnetic tweezers experiments consisted of a central fragment from the plasmid pBlueScriptIISK:pT181*cop608* flanked with digoxigenin or biotin-labeled DNA handles at its ends (Figure 6.3a). The central part was constructed from restriction of the plasmid pBlueScriptIISK:pT181*cop608* with BamHI and NotI (both from NEB), resulting in a final product of 7194 bp. The handles were PCR-generated from the plasmid pBlueScriptSK+ with appropriate oligos adding Dig-dUTP or Bio-dUTP (Roche) in the reaction and followed by restriction with NotI and BamHI, respectively. Subsequent ligation (T4 DNA ligase, NEB) with the central fragment resulted in an 8450 bp product, containing the RepC nicking site at 1480 bp from the biotinylated handle. The ligated product was then used in MT experiments. During the generation of the DNA substrates, we avoided the exposure of the substrates to intercalating agents as well as exposure to UV light. This procedure allowed the isolation of high levels of torsionally-constrained DNA molecules.

The nicking and religation assay of the TopIB enzyme (Figure 6.10) was performed with the same substrate employed in RepC experiments. The nicking assay with Nb.BbvCI was performed with a substrate based on the plasmid pSP73-JY0-BbvCI

[1]Purfication by Parvez Akhtar, PhD

employed in [Carrasco et al., 2013] of total length 5358 bp and with a single nicking site at 3106 bp from one end.

6.2.3 Magnetic tweezers assays

We employed the custom-built MT setup described in the section 3.2.1 of this thesis.

The integrity of the DNA molecules employed in this study was checked from the relation between extension and magnet turns at high and low forces. Single torsionally constrained DNA molecules display an asymmetrical response to magnet turns at high force with plectonemes formation only at positive turns, and a symmetrical curve at low forces where plectonemes are formed at both positive and negative turns (see section 3.6). From the linear parts of the graph, a change in extension as a function of magnet turns or linking number (ΔLk) could be calculated for a given force. For instance, at 0.34 pN, we obtained a value of 58 nm/turn. These constants allowed us to calculate the rate of supercoil removal, when required.

DNA relaxation measurements were done at the maximum data acquisition velocity of 500 Hz. The rest of experiments were performed at 60 Hz. All the experiments were done at room temperature in a buffer containing 100 mM KCl, 10 mM Tris-HCl, pH 8.0, 10 mM $Mg(CH_3COO)_2$, 10% ethylene glycol, 0.1% Tween-20 and 100 µg/ml BSA. Enzymes were injected into the fluid cell at 18 $\mu l \cdot min^{-1}$. Typical incubation times were 2-5 minutes.

6.2.4 Prediction of the atomic structure of RepC dimer

We have used Phyre2 software to obtain a prediction of the atomic structure for RepC based on homology modelling [Kelley et al., 2015]. Phyre2 is an online tool that employs an algorithm based on Hidden Markov models to generate a 3D structure of an input amino acid sequence using as template deposited structures on the PDB. The software returned a structure for the RepC monomer based on RepD (PDB entry 4CWC) having 81% of sequence identity corresponding to the central part of the protein [Carr et al., 2016]. The prediction was obtained with a 100% confidence match. The deposited RepD structure in the PDB is a dimer. In order to obtain a prediction of the dimeric structure for RepC, we aligned two RepC predicted monomers with the respective RepD monomers using PyMOL. Further analysis of the structures and the generation of the figures were carried out with PyMOL Molecular Graphics System (Version 1.8, Schrõdinger, LLC).

The statistical differences in the velocity of plectoneme release were analyzed by the bootstrap method [Efron and Tibshirani, 1994; Davison and Hinkley, 1997]. In brief, 10.000 artificial data sets are generated from random extraction with replacement from the original sample, generating new data sets with the same number of data in the original sample. This collection of samples was analyzed using a bootstrap-adapted Student's-t for non-equal variances (Welch test). The analysis was performed with a custom implementation of the algorithm described by Efron and Tibshirani [1994].

6.2.5 Corrections to nicking probablities from static RepC nicking experiments

In static RepC nicking experiments we cannot detect changes in extension during the incubation at low absolute values of ΔLk. Consequently, coilable molecules after the incubation time can be the result of a religation event after nicking instead of an absence of nicking activity. Thus, the probability of nicking is underrated if the experimental conditions do not allow for a direct observation of the nicking activty, i.e. during incubation at low $|\Delta Lk|$ before the DNA buckles. We can introduce a correction for this by obtaining the expected number of religation events, $\lambda_r = p_r N_n$, being N_n the number of nicked molecules and p_r the probability of religation obtained from experiments at high negative values of ΔLk, i.e. $\Delta Lk = -4$ and $\Delta Lk = -8$. We determined a religation probability of $p_r = 0.11$. Then, the corrected probability of nicking $\tilde{p}_{nick}$ can be calculated simply as $\tilde{p}_{nick} = (N_n + \lambda_r)/N$, where N is the total number of molecules.

6.3 Results

6.3.1 Nicking and religating activities of RepC protein

The double strand origin (*dso*) of the pT181 plasmid consists of 68 bps and includes both a specific RepC-binding sequence and its nicking site (Figure 6.2). The nicking site is within the gap sequence of the inverted repeat element II (IR-II), which is believed to fold to form a cruciform structure. The cruciform structure is not expected to spontaneously form because the IR-II region is GC-rich. However, this energy penalty to form the cruciform could be reduced upon RepC binding [Jin et al., 1997] or by untwisting the DNA. We developed a Magnetic Tweezers (MT) assay to study the dynamics of RepC nicking and religation activities on individual supercoiled DNA (Figure 6.3) [Koster et al., 2005; Strick et al., 1996]. In our assay, a single DNA molecule containing the *dso* that includes both RepC binding and nicking sites is tethered between a streptavidin coated magnetic bead (SA-bead) and the bottom glass surface of a flow cell (Figure 6.3a). Magnetic beads are manipulated by an external magnetic force produced by a couple of magnets placed above the flow cell. Vertical translation and rotation of the magnets about their axis induces stretching and torsion forces on the DNA molecule, respectively. Our MT assay allows us to supercoil a single torsionally-constrained (TC) DNA molecule by rotating the magnets along the optical axis. At low forces, positive or negative supercoils can be produced resulting in a reduction of DNA extension. At high forces, rotation of the magnets in the negative direction induces melting or denaturation of the DNA, while it is still possible to positively supercoil the DNA but with a higher number of magnet rotations (see section 3.6). In a typical measurement of RepC activity, a DNA molecule was negatively supercoiled by applying -30 turns (DNA untwisting) at a low force (0.34 pN) causing the corresponding decrease in extension (Figure 6.3b and Figure 6.3c). RepC protein was injected into the flow cell in a Mg^{2+}-free buffer. Then, the volume of the flow cell was exchanged with a Mg-containing buffer but without

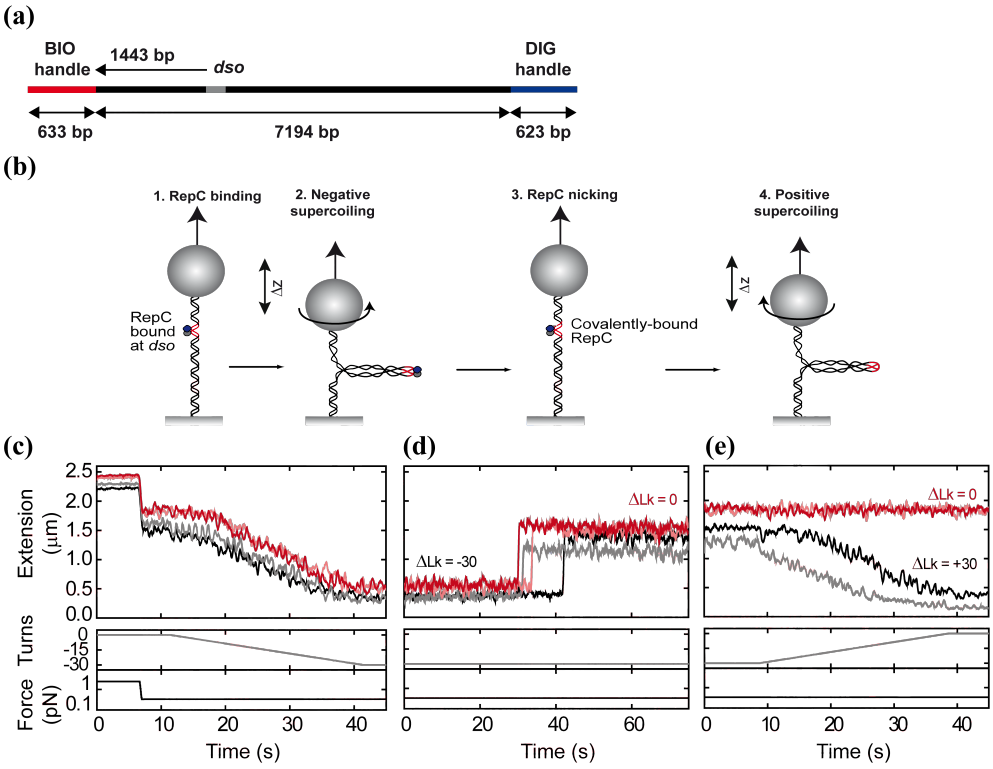

Figure 6.3: Nicking and religation activities of RepC probed by Magnetic Tweezers (MT). a) DNA substrate employed in MT assays. The substrate contains two handles labelled with biotins and digoxigenins that bind to the magnetic bead and glass surface, respectively. The central part contains the nicking site at 1480 bps from the biotinylated DNA end. b) MT assay. Torsionally-constrained (TC) DNA molecules are first negatively supercoiled by applying 30 negative rotations (steps 1 and 2). RepC nicks the DNA and a full recovery of the original height is observed (step 3). To check if DNA molecules can be further coiled, i.e., they have been religated, positive rotations at low force are applied (step 4). c) Extension, Turns, and Force time courses that correspond to the supercoiling of TC DNA molecules. d) Nicking activity of RepC. e) RepC religated a fraction of the nicked molecules as they could be further supercoiled by application of positive rotations. Data were acquired at 60 Hz and filtered down to 3 Hz (displayed).

protein. In most cases, we observed a sudden recovery of the initial height of the tethered DNA. This could be interpreted as the result of nicking activity by RepC at the *dso* and the subsequent release of supercoils (Figure 6.3b and Figure 6.3d) and suggests that the protein remained stably bound to the DNA during the buffer exchange process. The relaxation of supercoiled DNA always occurred in a single step. Control experiments performed with RepC in the absence of Mg^{2+} (data not shown) or with molecules lacking the *dso* sequence (Figure 6.4) did not result in relaxed DNA molecules, thus confirming that the nicking activity observed was indeed produced by RepC.

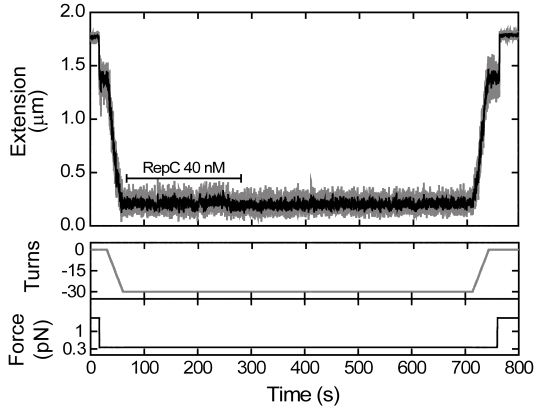

Figure 6.4: RepC cannot nick DNA without the *dso* of the pT181 plasmid. A torsionally-constrained (TC) DNA molecule of similar contour lenght ($\sim$ 6 kbp) without the *dso* of replication of the pT181 plasmid was negatively-supercoiled following injection of RepC protein. No nicking activity was detected for over 10 minutes. DNA remained TC since it was possible to recover the original extension by twisting the magnets back to zero rotations.

Interestingly, nicking was not concurrent for all molecules but occurred within the range of a few seconds and this could be due to the non-homogeneous arrival of Mg. Nicking rates in Y-junctions and plasmids measured by quenched-flow kinetics have been reported to be in the order of 1.5 and 25 s^{-1}, respectively [Arbore et al., 2012]. Finally, some DNA molecules could be further supercoiled demonstrating religation activity of RepC in the absence of torque and at an applied force of 0.34 pN (Figure 6.3e). However, the process of religation was inefficient as most of the DNA molecules remained nicked after several minutes of reaction. This could be due to a potential adverse effect of the rotation of the magnets with the ligation activity or simply to a inherent native low ligation efficiency of RepC.

6.3.2 RepC nicking activity is twist-dependent

The MT experiment described in Figure 6.3 was repeated with DNA containing positive supercoils. In this case, we did not observe any nicking event as DNA molecules remained supercoiled for several minutes. The inability of RepC to nick positively-supercoiled DNA allowed us to design experiments to probe the mechanistic detail of RepC nicking-religation activities. First, DNA molecules were positively supercoiled by applying +30 turns (anti-clockwise) and then RepC protein was injected into the fluid cell in a Mg-containing buffer. Turning the magnets from positive to negative rotations triggered the formation of negative supercoils and nicking by RepC. This is consistent with the protein remaining bound to the DNA during rotation of the magnet, as it was previously suggested in the buffer-exchange experiment (Figure 6.3). The nicking event can be detected as a sudden change of DNA extension up to its maximum, which remains unaltered regardless of the rotation of the magnet (Figure 6.5a). The rotation of the magnets is coupled to a change in the linking number

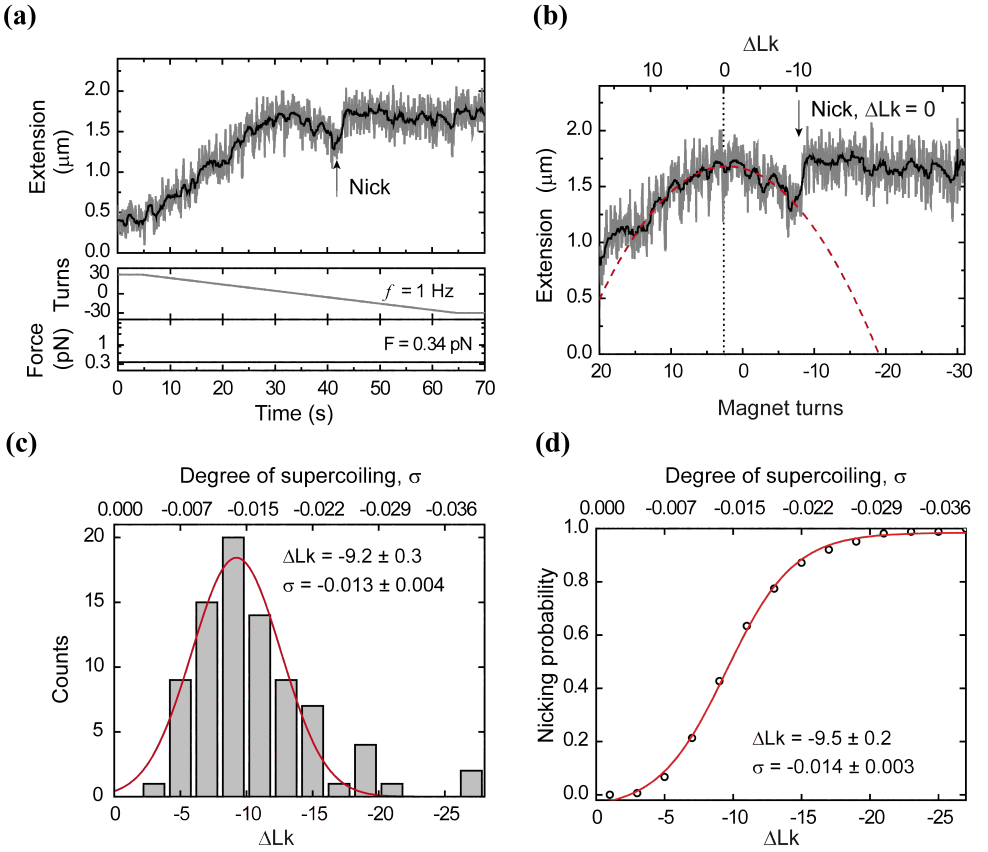

Figure 6.5: RepC nicking activity occurs only at negative rotations. a) DNA was nicked only at negative turns of the magnet. First, the DNA molecule was positively supercoiled (+30 rotations) and the protein injected in the cell. Then, the molecule was untwisted and negatively supercoiled by applying from +30 to -30 turns. Nick event detected (see arrow) at -7 turns. b) An offset between magnet turns and linking number was taken into account to determine the ΔLk at which RepC nicked the DNA. Data were acquired at 60 Hz and filtered down to 3 Hz (both displayed). Magnets were rotated at 1 Hz. Force was 0.34 pN, constant in the experiment. c) Histogram of ΔLk (N=83). The Gaussian fit provides a mean value of ΔLk and supercoiling degree σ (quoted in the figure). d) Normalized cumulative integral of the histogram representing the probability of nicking. Values of ΔLk and σ, quoted in the figure, relate to the characteristic ΔLk_c and σ_c at p = 0.5.

of the DNA (ΔLk). However, note that because of the presence of magnesium in the buffer with RepC, there is an offset of a few positive turns between ΔLk and magnet turns due to the change of supercoiling density of DNA caused by magnesium (Figure 6.5b). The effects on the topology of DNA due to divalent ions is well-characterized [Xu and Bremer, 1997; Rybenkov et al., 1997]. We therefore took into account these offsets for all the DNA molecules analyzed in this study. Both magnet turns and ΔLk vary equally irrespective of the initial offset produced by the presence of divalent ions, until the moment RepC nicks the DNA and the molecule is relaxed setting

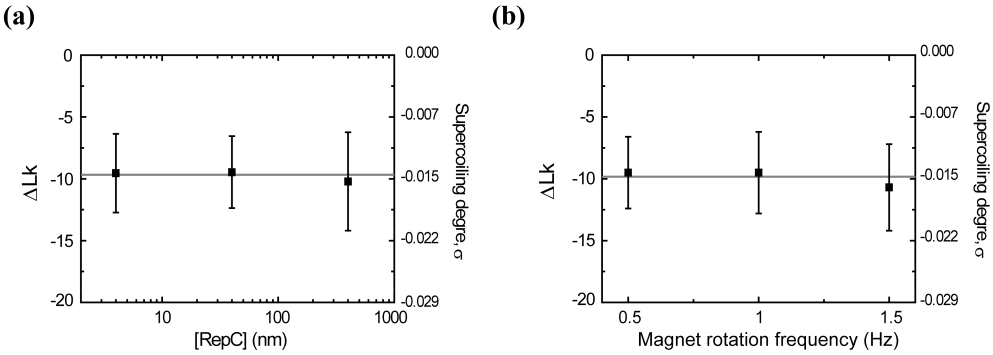

Figure 6.6: RepC nicking activity is independent of both RepC concentration and frequency of magnet rotation. a) Mean values and std errors of measurement of ΔLk of nicking for different concentrations of RepC. b) Mean values of ΔLk for RepC nicking obtained at 0.5, 1.0, and 1.5 Hz magnet rotation. Data obtained at 0.34 pN.

ΔLk to zero (Figure 6.3b). Beyond this point, turning of the magnets had no effect on ΔLk, which remains at zero. These data produced a distribution of ΔLk at the point where RepC nicks the DNA (Figure 6.5c). The distribution was fitted with a Gaussian function providing a mean value of ΔLk = -9.2 $\pm$ 0.3 (mean $\pm$ s.e.m) and a degree of supercoiling σ= -0.013 $\pm$ 0.004, calculated as $\Delta Lk/Lk_0$ where $Lk_0 = L/h$, with L the contour length of the DNA, and h the helical pitch (3.6 nm·turn^{-1}). Note that because our experimental procedure consists of progressively applying negative turns, the DNA is more likely to be nicked by RepC before reaching a large negative supercoiling value. The probability of nicking (p) is given by the normalized cumulative integral of the histogram (Figure 6.5d). This provided a characteristic ΔLk for $p = 0.5$ –from now on ΔLk_c– obtaining ΔLk_c = -9.5 $\pm$ 0.2. Since protein activity requires twisting of the DNA, our experiment suggests that RepC nicking is dependent on the formation of a particular structure in the DNA. Moreover, the fact that we only observed protein activity at negative turns, which will favor melting and formation of secondary structures in the DNA, strengthens the idea that a cruciform may be formed.

In order to exclude any effect of concentration in the reported critical linking number required for nicking we repeated the experiment described above for two additional concentrations. Average values of ΔLk were very similar and $\approx$ -10 (Figure 6.6a). We noticed, however, an increasing occurrence of nicking events at larger number of turns for the lower concentration used of 4 nM. This is consistent with the idea that binding of RepC to the substrate occurred during the course of the experiment in agreement with a limited availability of protein. Furthermore, at this low concentration, many DNA molecules were subjected to a much higher degree of negative supercoiling without nicking. We also explored if the frequency of twisting of the DNA may influence the probability of nicking occurring at a given force (Figure 6.6b). Turning the magnets at a lower (0.5 Hz) or higher (1.5 Hz) frequency had a small influence in ΔLk for p = 0.5. Efficient nicking occurred always at a value of

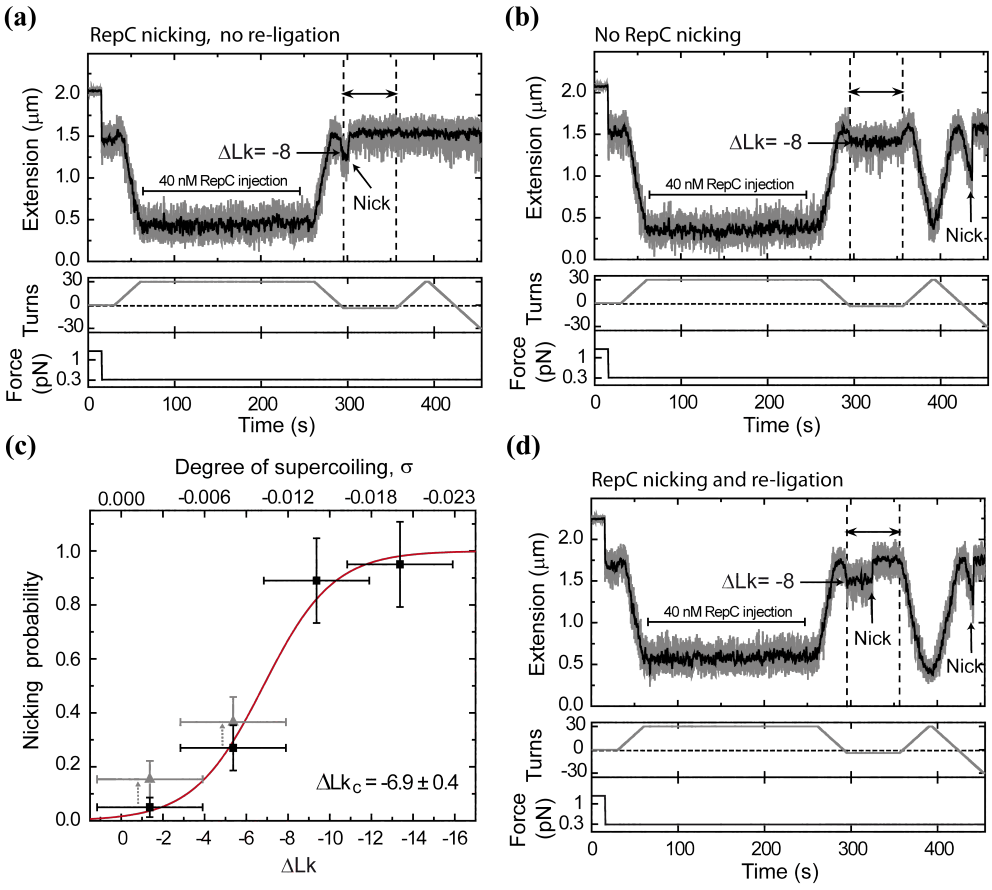

Figure 6.7: Static RepC nicking experiments performed on torsionally-constrained DNA (TC) set to a particular degree of twist. a) TC DNA molecules were first positively-supercoiled by applying +30 turns following injection of 40 nM RepC. Then, we applied turns to negatively supercoil the DNA up to an average ΔLk = -8. After one minute, we rotate again the magnets to check if the molecule is nicked or TC. The example in (a) represents the case where RepC nicked but did not religate the DNA. b) Case where RepC did not nicked the DNA over the one-minute incubation. c) Probability of nicking after one minute for different degrees of negative supercoiling (N=80). Grey symbols state for the probability of nicking considering religation events (see Materials and Methods). Probability error bars from counting error. ΔLk error bars determined from the std of the offset between magnet turns and ΔLk. d) Example where RepC nicked and religated the DNA. These events were rare as only 11% of the molecules were religated after one minute.

$\Delta Lk \approx$ -10 at the time scale of our experiments.

The dynamic assay reported in Figure 6.5 has the technical limitation that a minimum number of turns are required to detect a nicking event after the DNA buckles. The number of turns are coupled to a reduction in extension due to the formation of supercoils. Based on our noise level, sampling rate and measuring force (0.34 pN), we estimate that we minimally need to reduce the extension of the DNA by ∼85 nm

prior to observe a nicking event. Considering the number of turns before the DNA is buckled, we estimate that is necessary to apply an average of four turns to reduce the extension 85 nm. Below this number of turns it is not possible to distinguish if the molecule has been nicked or not. We therefore performed complementary static experiments where multiple TC DNA molecules were set to a particular degree of twist and exposed to RepC for one minute. Molecules that were nicked recovered the original extension (Figure 6.7a), in contrast to molecules that remained supercoiled, i.e. not nicked (Figure 6.7b). We also repeated these experiments at zero or near-zero turns. In these cases, we did not observe any reduction of extension, but we could determine if the molecule was nicked after one minute by simply turning the magnets again. We quantified the fraction of molecules that were nicked after one minute for different degrees of supercoiling (Figure 6.7c). As expected, none of the molecules were nicked if positively supercoiled, and almost none (2 events out of 40) were nicked at $\Delta Lk \approx 0$ (TC and relaxed). This measurement contrasts with the nicking rates reported in bulk using quenched-flow experiments and short linear dsDNA molecules ($\sim$1.5 s^{-1}) [Arbore et al., 2012]. Nevertheless, we should also bear in mind that in the previous work the experiments were performed with RepD –a sequence-similar protein– and low cross-reactivity has been reported between Rep proteins and the *dso*s of different plasmids [Thomas et al., 1990; Iordanescu, 1989]. The fraction of nicked molecules increased as more negative turns were applied following a very similar trend as reported in the continuous rotation experiment (Figure 6.5).

6.3.3 RepC nicking activity is force-dependent

Our data supports the view that a particular topological DNA structure is required for RepC nicking. This nicking was clearly sensitive to twist but it would also be expected to be sensitive to force because the extension-twist behavior of DNA is force-dependent (see section 3.6). We thus studied the nicking activity of RepC as a function of applied force following the dynamic assay of Figure 6.5. Because RepC only nicks under a negative turns regime, these experiments were restricted to a range of forces that allowed accumulation of writhe in the DNA (below 1 pN). We did experiments at forces between 0.05 and 1 pN and calculated the probability of nicking as a function of change in linking number. We observed a clear trend where more negative turns must be applied to maintain the probability of nicking, as applied force is reduced (Figure 6.8a). A linear fit of the data relative to the change in linking number at $p = 0.5$ and up to 0.34 pN (Figure 6.8b, red line) gave a value of about σ = -0.033 for a negatively-supercoiled plasmid at zero force. An increase of the force up to 0.34 pN implied a two-fold reduction in supercoiling degree. Beyond this force, we observed leveling of values due to the effects arising from melting in the DNA at negative turns at such forces. Melting of the DNA should in principle favor the formation of the cruciform, but on the other hand should also work against cruciform stabilization. Indeed, the presence of stable regions of melted DNA has been reported for forces above $\sim$0.5 pN [Strick et al., 1998]. We also observed that RepC cleavage

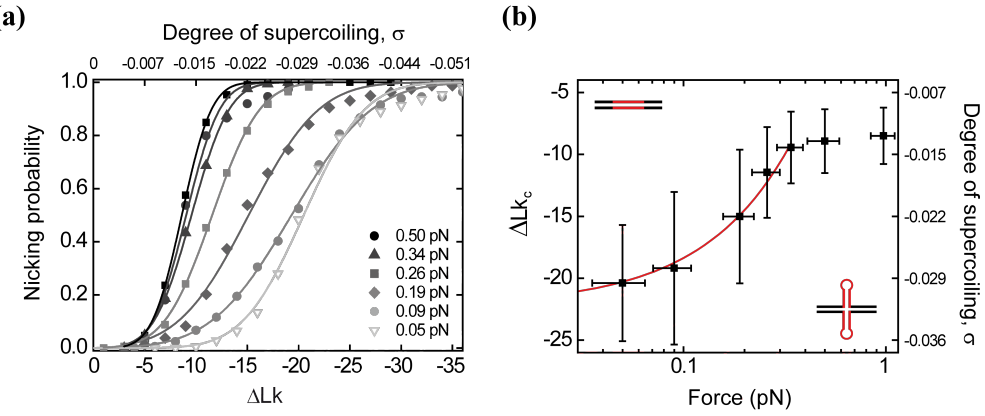

Figure 6.8: RepC nicking activity is force- and twist-dependent a) Nicking probability as a function of ΔLk for a range of forces of 0.05-1.0 pN. The nicking probability depends on the force applied to the DNA substrate, requiring a larger number of turns at low forces to observe RepC nicking. Also the distributions got sharper at higher forces. b) Characteristic ΔLk for $p = 0.5$ (ΔLk_c) at different forces. The line is a linear fit between 0.05 and 0.34 pN that provides a σ= -0.033 at zero force.

occurred within a range of turns that became narrower with increasing forces. RepC nicking at forces beyond 1 pN were still possible because we observed that the DNA could not be supercoiled again. However, these nicking events occurred before any reduction in extension, and thus it was not possible to determine the linking number at which RepC nicked the DNA. We expect a progressive reduction of the stability of the cruciform structure for forces above 1 pN until it is definitively inhibited. Opening of hairpins has been reported at forces between 10-14 pN, setting a limit for the formation of secondary structures [Essevaz-Roulet et al., 1997].

We next calculated the free energy (ΔG) of supercoiling using the expression

$$\Delta G(\sigma) = 10 N k_B T \sigma^2 \tag{6.1}$$

where N is the number of base pairs, k_B is the Boltzmann constant, and T is the absolute temperature [Vologodskii et al., 1990; Pulleyblank et al., 1975; Depew and Wang, 1975]. For σ = -0.033, we obtained a value of $\Delta G = 46$ kcal/mol. This value is in reasonable agreement with that reported for the formation of cruciform structures involving a small bubble intermediate, namely S-type cruciform structures, which is 40 kcal/mol [Lilley, 1989]. We also used the model by Marko [Marko, 2007; Lipfert et al., 2010, 2011; Moroz and Nelson, 1997] to calculate the change of free energy associated to DNA molecules at the supercoiling degree at which we obseved nicking, for intermediate forces of 0.19, 0.26, 0.34, and 0.5 pN. A detailed description of the procedures can be found in section 6.3.5. The results of this analysis were in reasonable agreement with that reported from the extrapolation at zero force and similar to the mentioned energy barrier proposed for a S-type cruciform.

Our experiments revealed that RepC nicking activity is twist- and force-dependent. The probability of getting a RepC-mediated DNA nicking event at a fixed

number of turns is higher at high forces than at low forces, at least for the range of forces studied in this work. For instance, a probability of nicking of 0.5 requires -10 turns at 0.35 pN and -20 turns at 0.05 pN. This force dependence on the probability of nicking is consistent with a model where the protein does not apply force on the DNA to form the required DNA structure for nicking. Otherwise, we would have observed a very limited effect of the force on the probability of nicking. Note that even at the very low force of 0.05 pN, the protein could not cleave the DNA unless a considerable number of negative turns were present. This idea contrasts with previously published work that proposed an active role of the protein in extruding the cruciform structure suitable for cleavage [Jin et al., 1997; Noirot et al., 1990].

6.3.4 RepC-DNA interactions during nicking and religation activities

The accepted mechanism of nicking-closing activity of RepC is that after the introduction of the nick, RepC remains covalently attached to the 5'-end of the nick via a phosphotyrosine bond [Thomas et al., 1990]. Measuring the velocity of DNA extension during supercoil release produced by RepC can provide insights into the nature of protein-DNA interactions that take place during nicking events [Koster et al., 2005]. Their experiments evidenced a reduction of the velocity of plectoneme release for nicking enzymes with high degrees of embracing of DNA than those having a puntual interaction with the DNA. Using a video camera capable of measuring at 500 fps, we were able to resolve in real-time the DNA relaxation curve during supercoil removal by RepC at 0.34 pN, which was characterized by a monotonic increase of DNA extension (Figure 6.9a). Data was fitted to a linear function to obtain the velocity of plectoneme release in ΔLk per second. The distribution of velocities was fitted to a Gaussian function (Figure 6.9b) showing a mean value for RepC of around 123 ± 3 $\Delta Lk{\cdot}s^{-1}$ (mean $\pm$ s.e.m.). We performed similar experiments with the human type-IB (hTopIB) topoisomerase (TopoGEN) that nicks the DNA following an identical chemical reaction than RepC, and with the nicking enzyme Nb.BbvCI (New England Biolabs). Experiments with the hTopIB gave a value of 110 $\pm$ 2 $\Delta Lk{\cdot}s^{-1}$ and the nicking enzyme a velocity of 136 $\pm$ 5 $\Delta Lk{\cdot}s^{-1}$, faster than both RepC and hTopIB. The statatistical significance of the difference between mean values were analyzed by boostraping (see section 6.2.3). We found p-values of 0.006, $6{\cdot}10^{-5}$ and 0.085 for the comparisons hTopIB-RepC, hTopIB-Nb.BbvC, and RepC-Nb.BbvC. These suggest a weak –but not null– interaction of the RepC active site with the DNA after nicking-induced plectoneme release.

Our experiments did not reveal large differences in plectoneme release velocities compared to those reported by Koster et al. [2005]. We explain the discrepancy with previous published data by the larger force employed in our case that increases the velocity of plectoneme release and may affect the interaction of the DNA with the protein. Still, our data consistently indicated an intermediate behaviour of RepC between the fastest response produced by the nicking enzyme and the slowest behaviour shown by the topoisomerase. We believe these differences in plectoneme

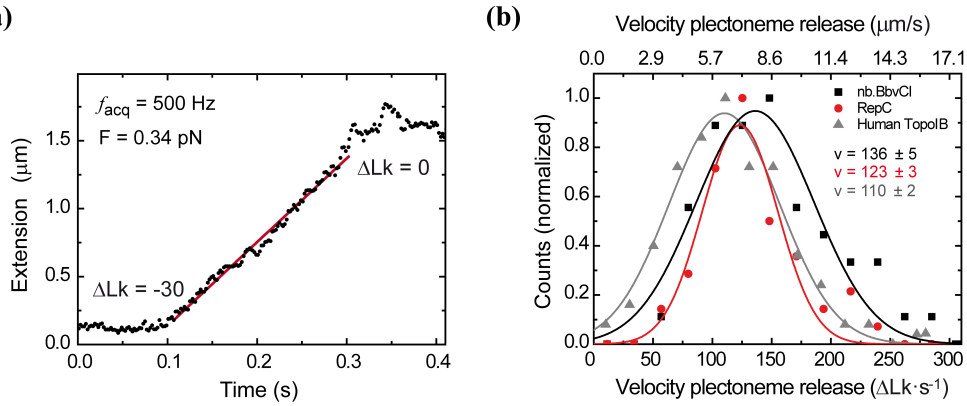

Figure 6.9: RepC-DNA interaction during plectoneme release. a) Characteristic supercoils removal time course taken at 0.34 pN. Approximately 30 turns were released in 0.3 s. Data were acquired at 500 Hz. b) Histogram of velocities of plectoneme release for RepC, hTopIB, and Nb.BbvCI. Frequencies are fitted to a Gaussian function and the mean and s.e.m values are quoted in the figure. RepC showed an intermediate velocity between the topoisomerase and the nicking enzyme.

release velocities reflect different types of protein-DNA interactions.

These differences between hTopIB and RepC became more evident in subsequent nicking-religation experiments (Figure 6.10). As mentioned above, RepC could only nick negatively supercoiled DNA containing the *dso* sequence (Figures 6.10a and 6.10b). Once nicked, we continuously rotated the magnet in the negative direction with the aim of negatively supercoiling the DNA again. If the protein seals the nick, we should expect to see a decrease in extension at the force employed in the assay (0.34 pN). RepC behaved as a very inefficient religating enzyme. Cycles of nicking and religation over the same molecule were rarely observed and very often the molecule was never religated back or required a long time for religation (several minutes). In the example shown in Figure 6.10a, RepC religated the DNA just once in about 6 minutes while it remained covalently linked to the DNA. However, other experiments did not show religation activity of RepC for tens of minutes. Long religation times for RepC have been described before in bulk biochemical assays [Arbore et al., 2012]. DNA supercoils were released by RepC protein always in a single step and this is consistent with the very low religating rate of the enzyme. In contrast, hTopIB could nick both positively and negatively supercoiled DNA and religate nicks very efficiently and for multiple cycles during the course of the experiment (Figures 6.10c and 6.10d). Occasionally, we observed a multistep relaxation curve as it has been previously reported (see arrow in Figure 6.10c) [Koster et al., 2005]. Differences in religation activities of hTopIB and RepC reveal a very different interaction with the DNA, and likely reflects their different specialized functions in the cell.

The poor religating activity of RepC was further confirmed in the assay performed in the absence of continuous magnet rotations (Figure 6.7). The example shown in Figure 6.7d is a positive religation case, but this was uncommon as only about

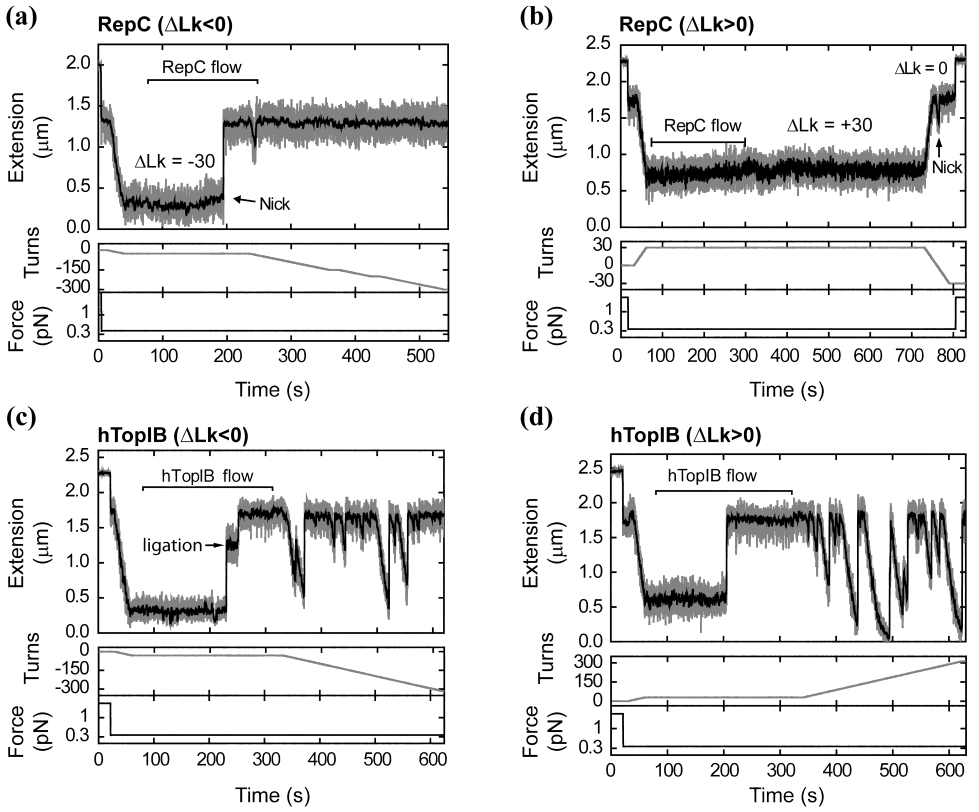

Figure 6.10: RepC-DNA interaction during nicking and religation activities. a) Nicking and religation experiment with RepC. RepC occasionally religated the nicked DNA. This is shown as a dip in the extension while the magnets continuously rotate in a negative direction. b) RepC cannot nick positively supercoiled DNA. c) and d) Nicking and religation experiment with hTopIB. hTopIB nicks and religates efficiently while the magnets continuously rotate in either negative or positive direction.

11% of the molecules were religated after one minute incubation at ΔLk = -8 turns and -4 turns. We analyzed the influence of religation events on the results obtained for the probability of nicking (see Materials and Methods) finding little influence on the global results (Figure 6.7c, grey symbols). The corrected probabilities $\tilde{p}$ are conditional probabilities dependent on religation and thus on the occurrence of a nicking event. Then, in absence of nicking during the incubation time, this analysis results in an overestimation of the real nicking probability. However, even under this highly conservative scenario, at $\Delta Lk \approx 0$ the probability of nicking yields a low value of $\tilde{p} = 0.15 \pm 0.05$.

The weak religation activity of RepC is consistent with its role in pT181 replication since after nicking it is not expected to religate the DNA immediately. Instead, after RepC-mediated nicking, plasmid replication is initiated, and only after the leading strand has been synthesized and the nicking site regenerated, RepC religates the plas-

mid (see Figure 6.1). Since there is no DNA replication in the experiments described here, RepC is able to religate the nicked DNA substrates at a very low efficiency.

6.3.5 Model for DNA supercoiling-dependent RepC nicking

In this section we will introduce a simple theoretical framework for the analysis of the data presented in Figure 6.8. Our observation is that as the stretching force on the DNA is reduced, a larger number of negative turns are required to detect RepC nicking. It has been proposed that the crossover where two DNA helices juxtapose can be a recognizable structure for supercoiling sensitive proteins [Zechiedrich and Osheroff, 1990]. We have discarded this possibility because at low forces plectonemes are formed with a lower number of turns and consequently nicking would occur earlier than at high forces, contrary to our observation. RepC nicking occurs in the plectonemic regime where the torque is constant. We hypothesize that the structure formed on the DNA needed for RepC nicking requires a characteristic supercoiling energy for its formation, because in the plectonemic regime the energy grows linearly with the number of turns. We have used Marko's model [Marko, 2007; Benham et al., 2009] to calculate the change of free energy associated to DNA molecules at the supercoiling degree at which we observe nicking, and at a particular stretching force F.

The work per contour length done on a supercoiled molecule with supercoiling density σ is given by $W(\sigma) = G(\sigma) - G(0)$, where G states for the supercoiling dependent free energy per contour length. The free energy per contour length at zero turns is

$$G(0) = -g(F) \tag{6.2}$$

where g corresponds to that of stretched and torsionally-relaxed DNA and can be approximated for forces between 0.2 and 10 pN as [Benham et al., 2009],

$$g(F) = F - \sqrt{\frac{k_B T F}{\xi}}, \tag{6.3}$$

where ξ is the bending persistence lenth of the DNA. In our experiments, nicking occurred in a region of coexistence of plectonemic and extended DNA, implying a reduction of extension with turns. In this region, the free energy per contour length is given by

$$G(\sigma) = \tau\omega_0|\sigma| - \frac{g}{1 - p_s/c_s} \tag{6.4}$$

where

$$\tau = \frac{1}{\omega_0}\sqrt{\frac{2pg(F)}{1 - p_s/c_s(F)}} \tag{6.5}$$

is the torque of the DNA in the coexistence state [Marko, 2007]. The parameter ω_0 is the contour-length rate of rotation of the relaxed double helix, $\omega_0 = 2\pi/(3.6$ nm) = 1.74 nm^{-1}. This parameter converts the supercoiling degree σ to an angle of

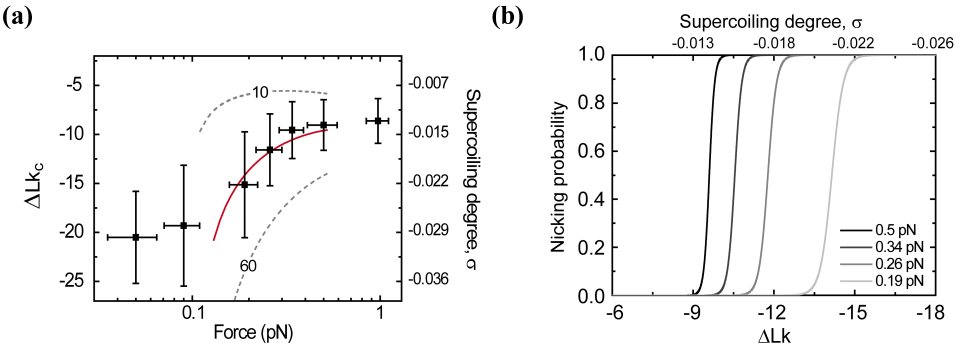

Figure 6.11: Model for RepC nicking activity. a) The data represented in Figure 6.8b was fitted to equation 6.8 and a $G^{\ddagger}$ = 33 kcal/mol is obtained from the best fit (red line). Dashed grey lines are guides to the reader for the expected ΔLk_c as a function of the force for $G^{\ddagger}$ = 10 and $G^{\ddagger}$ = 60 kcal/mol. b) Probabilities of nicking. The curves resemble the sigmoidal shape and the broadening of the transition at low forces observed in the experimental data.

rotation per contour length. The twist stiffness of the stretched state is defined by c_s, which dependency with the applied force is described by the Moroz and Nelson model [Moroz and Nelson, 1997]:

$$c_s(F) = k_B T \omega_0^2 C \left(1 - \frac{C}{4\xi}\sqrt{\frac{k_B T}{\xi F}}\right) \tag{6.6}$$

In this formula, C is twist persistence length. Analogously to c_s, the constant p_s states for the twist stiffness of the plectonemic state, $p_s = k_B T \omega_0^2 P$, being P the effective twist persistence length of the plectonemic DNA.

Combining equations 6.2 and 6.4, the work done at a certain supercoiling energy σ reads

$$W(\sigma) = \tau\omega_0|\sigma| - \frac{g}{1 - p_s/c_s} + g. \tag{6.7}$$

Note that here the energies are per contour length in force units. To obtain energy units, W have to be multiplied by the contour length of the DNA molecule ($L = 2.46$ μm for the 7194 bp substrate used in this work).

In order to determine the expected supercoiling degree at which RepC nicks, we have considered that a certain energy barrier $G^{\ddagger}$ have to be overcame to trigger the nicking activity of the protein. From equation 6.7, we solve for the value of σ such that $W(\sigma) = G^{\ddagger}$, which can be easily modified for ΔLk ($\Delta Lk = L\sigma h^{-1}$), obtaining:

$$\Delta Lk = \frac{1}{2\pi\tau}\left[G^{\ddagger} + Lg\left(\frac{1}{1 - p_s/c_s} - 1\right)\right] \tag{6.8}$$

In this expression the correction to energy units taking into account the contour length of the molecule L has been already included. In order to determine $G^{\ddagger}$ we fit the data to equation 6.8 where the following parameters are considered as fixed with

values described elsewhere: ξ = 45 nm [Herrero-Galán et al., 2013; Lipfert et al., 2014], C = 100 nm Mosconi et al. [2009]; Lipfert et al. [2011], and P = 15 nm [Brutzer et al., 2010; Kriegel et al., 2017]. Since g is an approximation for intermediate forces between 0.2 and 10 pN, and since we observe deviations at 0.5 pN and beyond due to melting we have used values between 0.18 and 0.5 pN for the fitting. The data are fitted to the model by numerical minimization of the χ^2 function of the experimental data and the theory as a function of $G^{\ddagger}$:

$$\chi^2(G^{\ddagger}) = \sum_{i=1}^{N} \left[\Delta Lk_c(F_i) - \Delta Lk_c^*(F_i, \Delta G^{\ddagger})\right]^2 \tag{6.9}$$

where ΔLk_c^* is the expected nicking position considering an energy barrier $G^{\ddagger}$ obtained from equation 6.8. From the best fit, we obtained a value for $G^{\ddagger} = 33$ kcal/-mol (Figure 6.11a). This value is in reasonable agreement with that reported from the extrapolation at zero force (46 kcal/mol) and similar to the energy barrier of $\sim$40 kcal/mol proposed for a S-type cruciform [Lilley, 1989].

Additionally, from the obtained energy barrier $G^{\ddagger}$ it is possible to characterize the expected probability of nicking p_n for a given ΔLk from a two states Boltzmann distribution as

$$p_n(\Delta Lk) = \frac{1}{1 + \exp\left(\frac{-(W(\Delta Lk) - G^{\ddagger})}{k_B T}\right)}. \tag{6.10}$$

The resulting probabilities as a function of force and ΔLk are represented in Figure 6.11b. Here, it has been considered that the energy before the DNA buckles is given by $G(\Delta Lk) = c_s \Delta Lk^2 h^2 / 2L$ [Marko, 2007]. The model resembles the experimental data shown in Figure 6.8a. This model assumes equilibrium conditions which is not the case in our experiments, as these are conducted under constant rotation of the magnets and, more importantly, the nicking reaction is not reversible. Nonetheless, our simple approximation qualitatively describes the general characteristics of the experimental data including the sigmoidal relation between ΔLk and the nicking probability as well as the sharpening of the transition when the force is increased.

6.4 Discussion

This work reports the first use of Magnetic Tweezers to directly measure RepC nicking at the single-molecule level. The advantage of MT over other single-molecule approaches is that a controlled force and/or twist can be applied to a single DNA molecule. Thus, information related to the degree of DNA supercoiling required for a specific enzymatic reaction can be obtained. In our assays, the extension of a ds-DNA molecule is followed in real-time and changes arising from the nicking of a supercoiled substrate can be detected with millisecond and nanometer resolution. Here, we show that the nicking activity of RepC is force- and twist-dependent suggesting that it requires particular torsional conditions for its specific activity. We also highlight a particular mode of interaction of RepC with the DNA which is different from that of the topoisomerase hTopIB that cleaves DNA using the same chemical mechanism.

Rep proteins were observed to nick and religate negatively supercoiled plasmid DNA containing the *dso* in vitro to form relaxed, covalently-closed molecules [Koepsel et al., 1985a]. This is consistent with the idea that formation of the nicking site for RepC is favored by the presence of negative supercoils, as reported in this study. However, the nicking activity of RepD, an homologous protein performing replication initiation in the staphylococcal plasmid pC221, has been observed in linear dsDNA substrates containing the *dso* although with a much lower nicking rate constant than for supercoiled plasmids [Arbore et al., 2012; Thomas et al., 1990; Chisty et al., 2013]. In our assay, we never observed nicking of positively supercoiled DNA and very rarely nicking of relaxed TC DNA subjected to a force as low as 0.34 pN.

Previous work has shown that the presence of a cruciform structure increases the efficiency of replication initiation [Noirot et al., 1990]. However, it is not fully understood if the cruciform structure is spontaneously formed under certain conditions of force and torque and RepC just binds and nicks the substrate, or if the Rep protein actively promotes formation of this structure upon binding and then nicks the substrate. Spontaneous formation of cruciform structures has been detected using MT as sudden changes in the extension of the DNA of hundreds of nanometers [Ramreddy et al., 2011]. However, we were unable to detect any change of extension attributable to the formation of the cruciform structure using similar forces, torques, and working conditions in our setup (not shown). This is very likely due to the different nucleotide sequence of the DNA substrate used in our study, as similar changes in extension as reported by Ramreddy et al. [2011] could be easily detected with our setup. Spontaneous formation of the cruciform structure at zero torque is very unlikely because the nick site of RepC in the pT181 DNA is at the tip of an inverted repeat element with much higher GC content than the rest of the plasmid. Consequently, it has been proposed that it is the binding of RepC that melts this region resulting in the extrusion of the inverted repeat element forming the cruciform structure needed for cleavage [Noirot et al., 1990; Jin and Novick, 2001]. However, we did not observe any change in extension that could be interpreted as the extrusion of the cruciform by RepC binding in a Mg-free buffer.

Instead, our data is consistent with a passive model (see below), where RepC binds at the *dso* but does not induce formation of the cruciform. Firstly, we found that at 0.34 pN RepC cleaves the DNA at a degree of supercoiling of $\Delta Lk \approx$ -10 turns, and that was largely insensitive to the speed of twisting of the DNA. This supports the idea that the structure suitable for cleavage is formed at a particular supercoiling degree. Secondly, we found that the probability of nicking is force-dependent. An active extrusion of the cruciform by the protein would have made RepC nicking activity largely insensitive to the range of the very low forces (0.05 - 1 pN) employed here, contrary to what we observed. Rather, RepC nicking activity was affected by the force and by the twist applied to the substrate. Our observation is that as the stretching force on the DNA is reduced, a larger number of negative turns are required to detect RepC nicking (Figure 6.8).

We report here the first precise measurement of the degree of DNA supercoiling

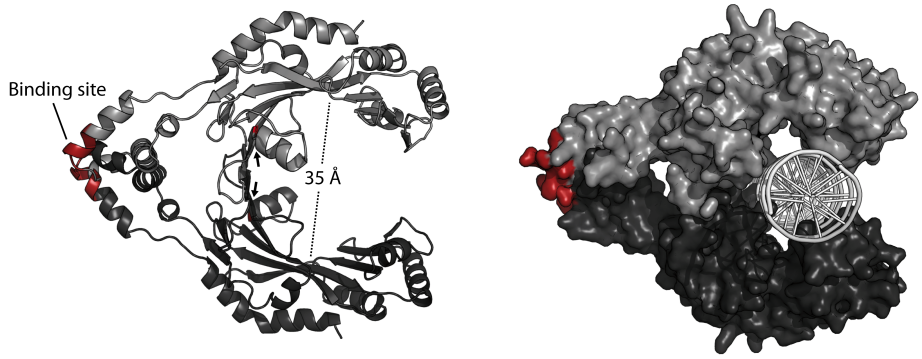

Figure 6.12: Structure prediction for RepC. Homology modeling of a structure for RepC based on its 80% sequence homology with RepD (PDB entry 4CWC). RepC is a dimer (monomers in dark and light grey) with separate DNA binding (red) and nicking domains (arrows). The dimer has a C-shape morphology with a cavity of 30-40 Å, enough to accommodate the DNA for cleavage (right).

necessary for DNA relaxation by Rep proteins. We determined a value of σ = -0.032 in the absence of force. Previous studies have made efforts to measure the degree of DNA supercoiling in vivo. It has been proposed that for *E. coli*, the average degree of supercoiling is $\sigma \approx$ -0.05 [Wang, 1984]. In contrast, Bliska and Cozzarelli [1987] showed that in *E. coli* the effective, purely torsional degree of supercoiling in the absence of protein stabilization was -0.025. Our measurement is therefore similar to the supercoiling degree found in vivo. This supports a model in which the cruciform structure forms passively due to the natural supercoiling of DNA, with RepC having only a minor contribution to its formation. We propose that the structure formed on the DNA needed for RepC nicking requires a characteristic supercoiling energy for its formation. We estimated a free energy associated with the degree of supercoiling required for nicking of 45 kcal/mol. This value is similar to the reported activation energies of S-type cruciform extrusion [Lilley, 1989]. S-type cruciforms are much more common than the C-type cruciform in which extrusion is dependent on a higher energy barrier (180-200 kcal/mol) [Lilley, 1989; Bikard et al., 2010]. The fact that the energy applied to the DNA to observe a nick by RepC coincides with the value reported for the extrusion of S-type cruciforms, further supports a passive model for RepC nicking activity. Once the critical degree of supercoiling is reached, the mechanism by which the protein recognizes the resulting secondary structure is not clear, and has been the subject of various studies [Gennaro et al., 1989; Noirot et al., 1990; Jin et al., 1997].

We have used the Phyre2 software to predict a structure for RepC [Kelley et al., 2015]. We found a very high confidence result based on the available structure of the closely-related RepD dimer (PDB entry 4CWC) [Carr et al., 2016]. The predicted structure is symmetric, despite RepC nicks only negative turns. This is not uncommon, as other chiral homodimer proteins have shown preference for cruciform struc-

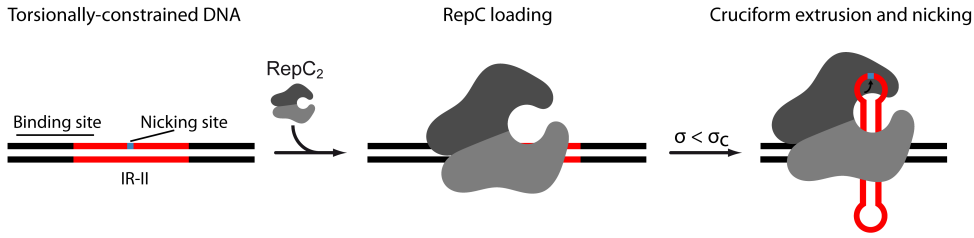

Figure 6.13: Model for the supercoiling-dependent nicking activity of RepC. RepC binds torsionally-constrained (TC) or torsionally-unconstrained DNA, but does not nick it because the nicking site is not available to the catalytic residue. Application of negative twist on TC DNAs promotes cruciform extrusion when a certain critical supercoiling densitiy is reached, placing the tip of the hairpin at the catalytic site of RepC resulting in efficient nicking.

tures or positively supercoiled DNA [Crisona et al., 2000; Zannis-Hadjopoulos et al., 2008]. RepC formed a C-shape dimer structure with a cavity of ∼30-40 Å internal diameter, where the catalytic tyrosine-191 responsible for DNA nicking is located. The C-shape predicted structure is compatible with the presence of friction that slows down the velocity of supercoil release and this is consistent with our measurements. Interestingly, residues 265-270 of RepC involved in its DNA binding activity [Dempsey et al., 1992] are located outside the cavity, on the convex side (Figure 6.12). Based on this structure and the experimental evidence reported here, we propose a model for RepC binding and nicking where both activities are independent. In unnicked DNA substrates –either circular DNA or linear DNA with constrained ends– RepC first binds and only at the appropriate supercoiling conditions, the cruciform is extruded, facilitating the contact of the nicking site with catalytic site of RepC. Linear DNA substrates with free ends may also be cleaved but that would require the sliding of one of the free ends of the DNA through the cavity where the catalytic site is located. This passive nicking model of RepC is compatible with previous observations that RepC can nick both linear dsDNA and ssDNA [Koepsel et al., 1986a; Koepsel and Khan, 1987] as well as other dsDNA structures with free ends [Arbore et al., 2012]. Linear substrates may just occasionally make contact with the active site resulting in cleavage. This should in principle result in much lower nicking rates than those found for plasmids as has been reported [Arbore et al., 2012], and supports the idea that nicking is favored by supercoiling. Indeed, plasmid DNA replication is initiated only when the DNA is supercoiled [Khan et al., 1981].

Supercoil-release experiments and nicking-religation assays with RepC and hTopIB revealed different interaction modes of these enzymes with the DNA, despite the fact that both nick and religate the DNA through the same chemical reaction. hTopIB could nick both positively and negatively supercoiled substrates and seal the nick with high efficiency [Champoux, 2001; Koster et al., 2005]. In contrast, RepC could only nick negatively supercoiled DNA and had very poor religation efficiency. These differences are likely a consequence of their distinct functions in the cell and we believe the cruciform structure is key to understanding their own mech-

anisms of functioning. In fact, because cruciform formation requires DNA melting, it will only form at negative turns, and this sets an additional filter for the selection of the replication initiator, increasing protein-DNA specificity. What happens with the cruciform structure when the DNA is nicked remains an open question, and the answer may explain the low efficiency of religation. Disruption of the cruciform by nicking may impede the correct placement of the 3'OH of the DNA at the catalytic domain of the Rep protein to seal the nick. The low efficiency of religation could be beneficial for the function of the enzyme that needs to load PcrA helicase to begin DNA unwinding during the process of RCR. A very fast religation rate would certainly hinder the helicase loading and replication initiation, as well as increase DNA-RepC protein exchange. Understanding at the single-molecule level how PcrA is loaded onto the DNA after RepC nicking and how this and the unwinding activities are affected by force and twist will be the subject of future work.

Our MT assay provides a basis to recapitulate at the single-molecule and in real-time the process of rolling-circle replication. Magnetic Tweezers allowed us to mimic the supercoiled state of plasmids and to investigate how this affects nicking and religation activities of replication initiator proteins involved in the first step of RCR. Our work highlights the relevance of the formation of the cruciform structure for RepC nicking, whose activity was dependent on the force and the twist applied to the DNA. We propose a passive model where RepC cleaves the appropriate structure when it is formed at a particular supercoiling energy imposed by force and twist. The implications of these mechanical requirements for nicking likely explain the low religation efficiency of this enzyme and provide a possible explanation for the differences observed with other nicking-religating enzymes such as type IB topoisomerases.

References

Arbore, C., Lewis, L. M., and Webb, M. R. Kinetic mechanism of initiation by RepD as a part of asymmetric, rolling circle plasmid unwinding. *Biochemistry*, 51(17):3684–93, 2012.

Benham, C. J., Harvey, S., Olson, W. K., de Witt, S. L., and Swigon, D. *Mathematics of DNA structure, function and Interactions*. Springer, 1st edition, 2009.

Bikard, D., Loot, C., Baharoglu, Z., and Mazel, D. Folded DNA in action: hairpin formation and biological functions in prokaryotes. *Microbiol. Mol. Biol. Rev.*, 74(4):570–88, 2010.

Bliska, J. B. and Cozzarelli, N. R. Use of site-specific recombination as a probe of DNA structure and metabolism in vivo. *J. Mol. Biol.*, 194(2):205–18, 1987.

Brutzer, H., Luzzietti, N., Klaue, D., and Seidel, R. Energetics at the DNA supercoiling transition. *Biophys. J.*, 98(7): 1267–76, 2010.

Carr, S. B., Phillips, S. E., and Thomas, C. D. Structures of replication initia-

tion proteins from staphylococcal antibiotic resistance plasmids reveal protein asymmetry and flexibility are necessary for replication. *Nucleic Acids Res.*, 44(5):2417–28, 2016.

Carrasco, C., Gilhooly, N. S., Dillingham, M. S., and Moreno-Herrero, F. On the mechanism of recombination hotspot scanning during double-stranded DNA break resection. *Proc Natl Acad Sci U.S.A.*, 110(28): E2562–2571, 2013.

Champoux, J. J. DNA topoisomerases: structure, function, and mechanism. *Annu. Rev. Biochem.*, 70:369–413, 2001.

Chang, T. L., Kramer, M. G., Ansari, R. A., and Khan, S. A. Role of individual monomers of a dimeric initiator protein in the initiation and termination of plasmid rolling circle replication. *J. Biol. Chem.*, 275(18):13529–34, 2000.

Chisty, L. T., Toseland, C. P., Fili, N., Mashanov, G. I., Dillingham, M. S., Molloy, J. E., and Webb, M. R. Monomeric PcrA helicase processively unwinds plasmid lengths of DNA in the presence of the initiator protein repd. *Nucleic Acids Res.*, 41(9):5010–23, 2013.

Crisona, N. J., Strick, T. R., Bensimon, D., Croquette, V., and Cozzarelli, N. R. Preferential relaxation of positively supercoiled DNA by E. coli topoisomerase IV in single-molecule and ensemble measurements. *Genes Dev.*, 14 (22):2881–92, 2000.

Davison, A. C. and Hinkley, D. V. *Bootstrap Methods and their Application.* Cambridge Series in Statistical and Probabilistic Mathematics, 1st edition, 1997.

del Solar, G., Giraldo, R., Ruiz-Echevarría, M. J., Espinosa, M., and Díaz-Orejas, R. Replication and control of circular bacterial plasmids. *Microbiol. Mol. Biol. Rev.*, 62(2):434–64, 1998.

Dempsey, L. A., Birch, P., and Khan, S. A. Six amino acids determine the sequence-specific DNA binding and replication specificity of the initiator proteins of the pT181 family. *J. Biol. Chem.*, 267(34):24538–43, 1992.

Depew, D. E. and Wang, J. C. Conformational fluctuations of DNA helix. *Proc. Natl. Acad. Sci. U.S.A.*, 72(11):4275–9, 1975.

Efron, B. and Tibshirani, R. J. *An Introduction to the Bootstrap.* Chapman & Hall/CRC Monographs on Statistics & Applied Probability, 1st edition, 1994.

Essevaz-Roulet, B., Bockelmann, U., and Heslot, F. Mechanical separation of the complementary strands of DNA. *Proc Natl Acad Sci U.S.A.*, 94(22):11935–40, 1997.

Gennaro, M. L., Iordanescu, S., Novick, R. P., Murray, R. W., Steck, T. R., and Khan, S. A. Functional organization of the plasmid pT181 replication origin. *J. Mol. Biol.*, 205(2):355–62, 1989.

Gruss, A. D., Ross, H. F., and Novick, R. P. Functional analysis of a palindromic sequence required for normal replication of several staphylococcal plasmids. *Proc Natl Acad Sci U.S.A.*, 84 (8):2165–69, 1987.

Herrero-Galán, E., Fuentes-Perez, M., Carrasco, C., Valpuesta, J., Carrascosa, J., Moreno-Herrero, F., and Arias-Gonzalez, J. Mechanical identities of

RNA and DNA double helices unveiled at the single-molecule level. *J. Am. Chem. Soc.*, 135(1):122–31, 2013.

Iordanescu, S. Specificity of the interactions between the Rep proteins and the origins of replication of staphylococcus aureus plasmids pT181 and pC221. *Mol. Gen. Genet.*, 217(2-3):481–7, 1989.

Jin, R. and Novick, R. P. Role of the double-strand origin cruciform in pT181 replication. *Plasmid*, 46(2):95–105, 2001.

Jin, R., Fernandez-Beros, M. E., and Novick, R. P. Why is the initiation nick site of an AT-rich rolling circle plasmid at the tip of a GC-rich cruciform? *EMBO J.*, 16(14):4456–66, 1997.

Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N., and Sternberg, M. J. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc*, 10(9):4456–66, 2015.

Khan, S., Carleton, S. M., and Novick, R. P. Replication of plasmid pT181 DNA in vitro: requirement for a plasmid-encoded product. *Proc. Natl. Acad. Sci. U.S.A.*, 78(8):4902–06, 1981.

Khan, S. A. Mechanism of replication and copy number control of plasmids in gram-positive bacteria. *Genet. Eng. (N.Y.)*, 15:183–201, 1996.

Khan, S. A. Rolling-circle replication of bacterial plasmids. *Microbiol. Mol. Biol. Rev.*, 61(4):442–55, 1997a.

Khan, S. A. Plasmid rolling-circle replication: highlights of two decades of research. *Plasmid*, 53(2):126–36, 1997b.

Khan, S. A. and Novick, R. P. Complete nucleotide sequence of pt181, a tetracycline-resistance plasmid from Staphylococcus aureus. *Plasmid*, 10(3): 251–59, 1983.

Koepsel, R. R. and Khan, S. A. Cleavage of single-stranded DNA by plasmid pT181-encoded RepC protein. *Nucleic Acids Res.*, 15(10):4085–97, 1987.

Koepsel, R. R., Murray, R. W., Rosenblum, W. D., and Khan, S. A. The replication initiator protein of plasmid pT181 has sequence-specific endonuclease and topoisomerase-like activities. *Proc. Natl. Acad. Sci. U.S.A.*, 82(20): 6845–9, 1985a.

Koepsel, R. R., Murray, R. W., Rosenblum, W. D., and Khan, S. A. Purification of pT181-encoded RepC protein required for the initiation of plasmid replication. *J. Biol. Chem.*, 260(14): 8571–77, 1985b.

Koepsel, R. R., Murray, R. W., and Khan, S. A. Sequence-specific interaction between the replication initiator protein of plasmid pt181 and its origin of replication. *Proc. Natl. Acad. Sci. U.S.A.*, 83 (15):5484–5488, 1986a.

Koepsel, R. R., Murray, R. W., Rosenblum, W. D., and Khan, S. A. Static and initiator protein-enhanced bending of DNA at a replication origin. *Science*, 233(4770):1316–18, 1986b.

Koster, D. A., Croquette, V., Dekker, C., Shuman, S., and Dekker, N. H. Friction and torque govern the relaxation of DNA supercoils by eukaryotic topoisomerase IB. *Nature*, 434(7033):671–74, 2005.

Kramer, M. G., Khan, S. A., and Espinosa, M. Plasmid rolling circle replication: identification of the RNA polymerase-directed primer RNA and requirement for DNA polymerase I for lagging strand synthesis. *EMBO J.*, 16 (18):5784–95, 1985.

Kriegel, F., Ermann, N., Forbes, R., Dulin, D., Dekker, N., and Lipfert, J. Probing the salt dependence of the torsional stiffness of DNA by multiplexed magnetic torque tweezers. *Nucleic Acids Res.*, Epub ahead of print, 2017.

Lilley, D. M. J. Structural isomerization in DNA: The formation of cruciform structures in supercoiled DNA molecules. *Chem. Soc. Rev.*, 18:53–83, 1989.

Lipfert, J., Kerssemakers, J. W., Jager, T., and Dekker, N. H. Magnetic torque tweezers: measuring torsional stiffness in DNA and RecA-DNA filaments. *Nat. Methods.*, 7(12):977–80, 2010.

Lipfert, J., Wiggin, M., Kerssemakers, J. W., Pedaci, F., and H., D. N. Freely orbiting magnetic tweezers to directly monitor changes in the twist of nucleic acids. *Nat. Commun.*, 2:439, 2011.

Lipfert, J., Skinner, G. M., Keegstra, J. M., Hensgens, T., Jager, T., Dulin, D., Kober, M., Yu, Z., Donkers, S. P., Chou, F. C., Das, R., and Dekker, N. H. Double-stranded RNA under force and torque: similarities to and striking differences from double-stranded DNA. *Proc. Natl. Acad. Sci. U.S.A.*, 111(43): 15408–13, 2014.

Marko, J. F. Torque and dynamics of linking number relaxation in stretched supercoiled DNA. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.*, 76:021926Ű39, 2007.

Moroz, J. D. and Nelson, P. Torsional directed walks, entropic elasticity, and DNA twist stiffness. *Proc. Natl. Acad. Sci. U.S.A.*, 94(26):14418–22, 1997.

Mosconi, F., Allemand, J., Bensimon, D., and Croquette, V. Measurement of the torque on a single stretched and twisted DNA using magnetic tweezers. *Phys. Rev. Lett.*, 102(7):078301, 2009.

Noirot, P., Bargonetti, J., and Novick, R. P. Initiation of rolling-circle replication in pT181 plasmid: initiator protein enhances cruciform extrusion at the origin. *Proc. Natl. Acad. Sci. U.S.A.*, 87(21):8560–4, 1990.

Pulleyblank, D. E., Shure, M., Tang, D., Vinograd, J., and Vosberg, H. P. Action of nicking-closing enzyme on supercoiled and nonsupercoiled closed circular DNA: formation of a Boltzmann distribution of topological isomers. *Proc. Natl. Acad. Sci. U.S.A.*, 72(11):4280–4, 1975.

Ramreddy, T., Sachidanandam, R., and Strick, T. R. Real-time detection of cruciform extrusion by single-molecule DNA nanomanipulation. *Nucleic Acids Res.*, 39(10):4275–83, 2011.

Rasooly, A. and Novick, R. P. Replication-specific inactivation of the pT181 plasmid initiator protein. *Science*, 262(5136):1048–50, 1993.

Redinbo, M. R., Stewart, L., Kuhn, P., Champoux, J. J., and Hol, W. G. Crystal structures of human topoisomerase I in covalent and noncovalent complexes

with DNA. *Science*, 279(5356):1504–13, 1998.

Ruiz-Masó, J. A., Machón, C., Bordanaba-Ruiseco, L., Espinosa, M., Coll, M., and Del Solar, G. Plasmid rolling-circle replication. *Microbiol. Spectr.*, 3(1), 2015.

Rybenkov, V. V., Vologodskii, A. V., and Cozzarelli, N. R. The effect of ionic conditions on DNA helical repeat, effective diameter and free energy of supercoiling. *Nucleic Acids Res.*, 25(7):1412–18, 1997.

Seidel, R., van Noort, J., van der Scheer, C., Bloom, J., Dekker, N. H., Dutta, C. F., Blundell, A., Robinson, T., Firman, K., and Dekker, C. Real-time observation of DNA translocation by the type I restriction modification enzyme ecoR124I. *Nat. Struct. Mol. Biol.*, 11(9): 838–43, 2004.

Strick, T. R., Allemand, J. F., Bensimon, D., Bensimon, A., and Croquette, V. The elasticity of a single supercoiled DNA molecule. *Science*, 271(5257):1835–37, 1996.

Strick, T. R., Allemand, J. F., Bensimon, D., and Croquette, V. Behavior of supercoiled DNA. *Biophys. J.*, 74(4):2016–28, 1998.

Thomas, C. D., Balson, D. F., and Shaw, W. V. In vitro studies of the initiation of staphylococcal plasmid replication. specificity of RepD for its origin (oriD) and characterization of the Rep-ori tyrosyl ester intermediate. *J. Biol. Chem.*, 265(21):5519–30, 1990.

Vologodskii, A. V., Lukashin, A. V., Anshelevich, V. V., and Frank-Kamenetskii, M. D. Fluctuations in superhelical DNA. *Nucleic Acids Res.*, 6(21):967–82, 1990.

Wang, J. C. DNA supercoiling and its effects on the structure of DNA. *J. Cell. Sci. Suppl.*, 1:21–9, 1984.

Xu, Y. C. and Bremer, H. Winding of the DNA helix by divalent metal ions. *Nucleic Acids Res.*, 25(20):4067–71, 1997.

Zannis-Hadjopoulos, M., Yahyaoui, W., and Callejo, M. 14-3-3 cruciform-binding proteins as regulators of eukaryotic DNA replication. *Trends Biochem. Sci.*, 33(1):44–50, 2008.

Zechiedrich, E. L. and Osheroff, N. Eukaryotic topoisomerases recognize nucleic acid topology by preferentially interacting with DNA crossovers. *EMBO J.*, 9(13):4555–62, 1990.

Part IV

Conclusions and future perspectives

7

Conclusions and future perspectives

This dissertation belongs to the field of single molecule biophysics. We have employed magnetic tweezers (MT) to study the interaction between bacterial proteins and single DNA molecules. A detailed description of the MT apparatus, including operation tips and state-of-the art improvements have been included as an introductory section (Chapter 3).

The results obtained from two independent biological systems involved in chromosome organization and DNA plasmid replication are shown in the research projects corresponding to Parts II and Part III of this thesis. In the following sections, the main conclusions of each project are summarized. Based on the conclusions obtained, additional confirmations required to support our models as well as open questions are proposed as future perspectives.

7.1 The role of ParB in chromosome condensation

This project explored the interaction between the protein ParB from *B. subtilis* and the DNA. The ParABS system is involved in the formation of a *organization center* responsible of chromosome segregation. The description of ParB as a loader of SMC proteins to the DNA, have increased the interest of the community on the ParB-*parS* system. The SMC complex has been described as one of the main actors in the organization of the bacterial chromosome, but there are poor evidences of its ability to bind DNA. This lead us to investigate the mechanisms that could facilitate the interaction between the SMC complex and the DNA as a result of ParB binding to DNA.

7.1.1 Conclusions

We find a novel DNA condensation activity for ParB at forces below 2 pN (Chapter 4). Furthermore, ParB stabilizes plectonemes in torsionally-constrained molecules and braiding interactions between DNA substrates in double-tethered beads. These phenomena are explained by the formation of protein-protein interactions between remote DNA regions.

Strikingly, ParB-dependent DNA condensation is still present in absence of the *parS* sequence, showing similar condensation parameters to that observed for *parS*-containing substrates. Moreover, Freely Orbiting MT experiments reveal that ParB binding and condensation of *parS*-containing DNAs are not linked to an organized

structural rearrangement of the DNA. Together, these results indicate that the *parS* sequence is not a *nucleation site* for ParB binding and condensation.

In the second part of this project (Chapter 5), we analyzed the implications of the ParB domains in DNA binding and condensation. Biochemical experiments evidenced the presence of a novel DNA binding domain located at the C-terminal domain (CTD) of ParB. Mutations impairing DNA binding by the CTD strongly abrogate DNA condensation. In addition, free CTD prevented DNA condensation when co-incbuated with wild-type ParB. The ability of the CTD to inhibit ParB-dependent condensation is retained even when CTD mutants unable to bind DNA are used. This indicates that condensation is dependent on both CTD binding to DNA and CTD dimerisation. However, a single mutation at the N-terminal domain (NTD) strongly depletes the efficiency of condensation. This indicates that ParB condensation is not simply the result of ParB dimerization between distant DNA regions.

The main conclusions of this part are summarized in the following points:

- ParB non-specific binding and ParB-ParB interactions result in DNA condensation. Neither non-specific DNA binding nor DNA condensation are dependent on the *parS* sequence. Consistently with this observation, our results indicate a random non-specific binding of ParB to DNA, in contrast to the previously proposed model of a well-ordered filament spreading from *parS* sequences.

- DNA binding and dimerization of ParB by the C-terminal domain are necessary for DNA condensation. Nonetheless, a single mutation at the N-terminal domain not affecting DNA binding, efficiently impairs ParB-dependent DNA condensation. These results illustrate the complex architecture of the ParB protein and the interplay between different domains to condense the DNA.

7.1.2 Future perspectives

Magnetic tweezers experiments revealed a novel DNA condensation activity of the protein ParB and suggested a dependency on dimerization through the CTD for this activity. Mutations at the dimerization patch of the CTD are a necessary control to reinforce current results. These mutations result in unfolded proteins and the production and purification of stable proteins and their characterization by MT will be the subject of future work in our group.

Additional experiments are required to understand the mechanism of DNA condensation by ParB. The inhibition of ParB condensation in the presence of a large excess of CTD (Chapter 5) can be explained by two different mechanisms. In a first hypothesis, the formation of wild-type ParB and CTD heteroligomers (ParB-CTD) can disrupt the interaction of the wild-type protein with the DNA. In a second hypothesis, ParB condensation is a consequence of entrapping DNA by oligomerization between by the N-terminal domain (NTD) and the CTD of ParB. In this scenario, the absence of the NTD in the CTD truncation abolishes the formation of ParB com-

plexes, and consequently the driving force of the condensation, without affecting the binding to DNA.

In order to test these hypotheses we have started experiments with fluorescence versions of both ParB and CTD using a combined MT-TIRF setup. The combination of MT with TIRF microscopy allows us to monitor protein binding while preventing DNA condensation. Our preliminary experiments suggest a decoration of the CTD along ParB-coated DNAs, hence supporting the second hypothesis. This scenario is also favoured by the absence of condensation reported on NTD mutants (R80A). More research is still necessary to determine the *minimal condensation unit*. For this objective it is convenient to design an experimental protocol to observe the formation of ParB-dependent DNA loops as a result of the interaction between a couple of ParB protomers. The use of DNA substrates with two separated *parS* sites is the more obvious approximation. However, the low specificity for the *parS* sequence requires a tight tuning of the experimental conditions. Again, the combination of MT with TIRF microscopy is a powerful tool to investigate the specific binding of ParB to *parS* sequences at the single-molecule level. The strict localisation of ParB around *parS* sequences found *in vivo* contrasts with the low specificity found in both ensemble and single-molecule experiments. Therefore, these experiments are critical to understand this apparent discrepancy.

Our mid term goal is to explore the role of ParB binding in the loading of the SMC complex. It has been described that the interaction with ParB bound to *parS* sequences is needed for the binding of SMC proteins to the DNA. Since SMC proteins are key elements of chromosome condensation and topological organization, their lack of DNA binding in absence of ParB is a striking feature of their activity. A plausible explanation is that the particular structure adopted by condensed DNA around the *parS* sequence is required for SMC binding. There are a manifold of proposed models for how the SMC complexes work to condense the chromosome and currently this is a very active and competitive research topic. Our single-molecule MT experiments are an invaluable tool to study the mechanisms of the ParB-dependent activity of SMC proteins on the basis of the previously acquired knowledge on ParB activity.

7.2 Force and twist dependence of RepC nicking activity

In the so-called rolling circle replication mechanism, replication initiator proteins (Rep) are involved in the generation of a nick at the origin of replication of plasmids. This promotes the loading of the replication machinery and the synthesis of DNA proceeds from the nicking site. The mechanistic details of this process are largely unknown albeit the important implication of plasmids for the biology of Bacteria. In this project we have studied the nicking activity of the replication initiator protein RepC and its dependency with the supercoiling state of the DNA.

7.2.1 Conclusions

We have designed an experimental protocol to measure the supercoiling-dependent activity of RepC (Chapter 6). We find that the incubation of RepCwith negatively-supercoiled DNA results in the release of up to 30 turns in a single step. In contrast, we never observe nicking activity in positively supercoiled DNA. We have exploited this characteristic to precisely measure the degree of negative supercoiling necessary for RepC nicking. RepC is incubated in positively-supercoiled DNA before the application of negative turns to promote the nicking activity of the enzyme. We measured the supercoiling degree leading to RepC nicking, finding that this is dependent on the applied forces. Somewhat surprising, the higher the force the lower the number of supercoils triggering RepC nicking, at least for the range of forces explored in this work (0.05 pN to 1 pN). From the analysis of these data, we have obtained supercoiling energies compatible to the activation energies reported for the extrusion of DNA cruciforms. The presence of sequences potentially resulting in DNA cruciforms near the RepC binding site has been described previously by other groups. Our results support a model where the activity of the replication initiator protein relies on the extrusion of the cruciform.

We also investigated the nature of the interaction between the DNA and RepC, both during and after the nicking reaction. We compared the activity of RepC with that of the human topoisomerase IB (TopoIB). This enzyme relaxes torsional stress on the DNA following a chemical mechanism analogous to RepC. However, we find that TopoIB behaves very differently to RepC. While TopoIB shows a highly efficient DNA religation activity, we rarely detected religation events for RepC. Furthermore, the activity of TopoIB is independent on the supercoiling state of the DNA, contrary to our observation for RepC. This indicates that while both enzymes have a common catalytic mechanism, the interaction mechanism with the DNA should differ between both enzymes. Combining our single-molecule experiments with homology modelling, we conclude that RepC is a dimer with a C-shaped structure.

We propose a simple mechanistic model, where DNA supercoiling triggers the formation of a cruciform structure. The extrusion of the cruciform places the DNA nicking site at the catalytic residue of RepC. Our results suggest spontaneous extrusion of the cruciform when the appropriate conditions of supercoiling and force are satisfied, in contrast to previously-proposed models attributing an active role for RepC in cruciform extrusion.

The main conclusions of this project are summarized in the following points:

- The activity of RepC is dependent on the supercoiling state of the DNA. We found that negative supercoiling is necessary for RepC nicking activity. Interestingly, while RepC nicking activity is highly efficient on negatively-supercoiled DNA, our results indicate a low religation activity. These results contrast with the experiments performed with topoisomerase enzymes whose nicking mechanism is similar to that of RepC.

- The activity of RepC is dependent on the stretching force on the DNA. Our results indicate the need of an increase in the number of negative superturns for the activity of RepC at low forces compared to higher forces. The supercoiling energies are consistent with the formation of a cruciform DNA structure.

- Magnetic tweezers experiments as well as homology modelling indicate a C-shaped structure for the RepC dimer. We propose a working model for RepC where cruciform extrusion places the nicking position on the DNA at the active residue of the protein.

7.2.2 Future perspectives

Our observations are consistent with the dependency of RepC activity on the formation of a cruciform structure. However, in our hands we have not been able to resolve the formation of these structures. Therefore, it cannot be discarded the possibility of transitory structures or melting bubbles as sufficient for RepC nicking activity. In order to understand the implications of secondary DNA structures on the activity of RepC, it would be useful to perform experiments on DNA substrates lacking on the cruciform sequence where the GC-content is mantained.

In addition, the use of shorter molecules would be convenient to resolve changes in extension attributed to the extrusion of a cruciform or alternative secondary structures. Tuning the conditions to resolve the formation of a cruciform structure at the single-molecule level will provide an unvaluable tool to directly quantify the implications of these structures on the binding and nicking activity of RepC.

The incorporation of the rest of the proteins involved in the rolling-circle replication is a step forward to the understanding of the reaction as a whole. The subsequent enzyme involved in the reaction after RepC nicking is the PcrA helicase, whose activity is in fact highly dependent on the activity of RepC. PcrA is loaded on the origin of replication by RepC and unwinds the DNA duplex proceeding from the RepC nicking site. The production of single-stranded DNA (ssDNA) can be monitored by magnetic tweezers since at the applied forces in this work, the different mechanical properties of ssDNA compared with double stranded DNA results in a shortening of the tether. Hence, it is possible to measure at the single-molecule level the activity of the helicase PcrA in complex with RepC. The experimental conditions have been optimized and interesting results have been obtained by Dr. Carolina Carrasco. The characterization of the activity of PcrA as well as its implications for RepC-nicking activity will be the subject of a future publication.

8

Conclusiones y perspectivas futuras

Esta disertación pertenece al campo de la biofísica de molécula única. Hemos empleado pinzas magnéticas (PM) para estudiar la interacción entre proteínas bacterianas y moléculas únicas de ADN. En el Capítulo 3 se ha incluido una descripción detallada del sistema de PM, con consejos de operación y mejoras de vanguardia.

Los resultados obtenidos de dos sistemas biológicos independientes involucrados en la organización del cromosoma y en la replicación del ADN plasmídico se muestran en los proyectos de investigación correspondientes a la Parte II y Parte III de esta tesis. En las siguientes secciones de este capítulo, se resumen las conclusiones de cada proyecto. Basándonos en dichas conclusiones, se plantean confirmaciones requeridas para apoyar nuestros modelos así como preguntas abiertas como perspectivas futuras.

8.1 El papel de ParB en la condensación del cromosoma

En este proyecto se ha explorado la interacción entre la proteína ParB de *B. subtilis* y el ADN. El sistema ParABS está involucrado en la formación de un *centro de organización* responsable de la segregación del cromosoma. La descripción de ParB como cargador de las proteínas SMC al ADN ha incrementando el interés de la comunidad en el sistema ParB-*parS*. El complejo SMC se ha descrito como uno de los actores principales en la organización del cromosoma bacteriano, pero hay pocas evidencias acerca de su capacidad para unirse al ADN. Esto nos llevó a investigar los mecanismos que podrían facilitar la interacción entre el complejo SMC y el ADN como resultado de la unión de ParB al ADN.

8.1.1 Conclusiones

Encontramos que ParB puede condensar moléculas de ADN a fuerzas menores de 2 pN (Capítulo 4). Esta actividad no había sido descrita previamente. Además, ParB estabiliza plectonemas en moléculas torsionalmente constreñidas e interacciones en ADN trenzado esferas ancladas por más de un ADN. Estos fenómenos pueden ser explicados considerando interacciones proteína-proteína entre regiones distantes del ADN.

Sorprendentemente, la condensación de ParB es independiente de la secuencia *parS*, obteniendo parámetros de condensación similares en presencia y ausencia de dicha secuencia. Además, experimentos de PM Libre-Orbitantes, revelan que la unión

de ParB y la condensación del ADN no están asociadas a cambios estructurales organizados del ADN. En conjunto, estos resultados indican que la secuencia *parS* no es un *sitio de nucleación* para la unión y condensación de la proteína ParB.

En una segunda parte de este proyecto (Capítulo 5) analizamos las implicaciones de los dominios de ParB en la unión a ADN y la condensación. Experimentos bioquímicos evidencian la presencia de un nuevo dominio de unión a ADN localizado en el dominio C-terminal (DCT) de ParB. Las mutaciones que inhiben su unión a ADN por el DCT no muestran actividad condensadora del ADN. Además, el DCT previene la condensación del ADN cuando se co-incuba con ParB salvaje. La capacidad del DCT para inhibir la condensación del ADN por ParB no se ve afectada cuando se usan mutantes del DCT libre incapaces de unirse al ADN. Esto indica que la condensación es dependiente de la unión a ADN por el DCT así como de la dimerización por dicho dominio. Sin embargo, una única mutación en el dominio N-terminal (DNT) decrementa fuertemente la eficiencia de la condensación. Esto indica que la condensación del ADN por ParB no es simplemente el resultado de la dimerización de ParB entre regiones remotas del ADN.

Las conclusiones principales de esta parte se resumen en los siguientes puntos:

- La unión no específica de ParB al ADN y las interacciones ParB-ParB dan lugar a condensación del ADN. Ni la unión no específica ni la condensación dependen de la secuencia *parS*. De acuerdo con esta observación, nuestros resultados indican una unión no específica aleatoria, contrariamente al modelo previamente propuesto, donde la unión no específica era el resultado de la formación de un filamento a partir de la secuencia *parS*.

- La unión de ParB a ADN y la dimerización por el dominio C-terminal son necesarios para la condensación del ADN. No obstante, una mutación puntual en el dominio N-terminal que no afecta a la unión a ADN inhibe eficientemente la condensación del ADN dependiente de ParB. Estos resultados ilustran la compleja arquitectura de la proteína ParB y las cooperatividad entre los diferentes dominios para condensar el ADN

8.1.2 Perspectivas futuras

Nuestros experimentos de pinzas magnéticas revelan una nueva actividad condensadora de la proteína ParB, y sugieren que esta actividad depende de la dimerización por el DCT. Con el fin de reforzar nuestros resultados será necesario el estudio control con mutantes en la región de dimerización del DCT. Estas mutaciones dan lugar a proteínas mal plegadas y la purificación de proteínas estables así como su caracterización con pinzas magnéticas, será sujeto de estudio por nuestro grupo.

También son necesarios experimentos adicionales para entender los mecanismos de condensación de ParB. La inhibición de la condensación de ParB por el DCT (Capítulo 5) se puede explicar por dos mecanismos distintos. En una primera hipótesis, la formación de heteroligomeros entre ParB salvaje y el DCT (ParB-DCT) podría in-

hibir la interacción de la proteína salvaje con el ADN. En una segunda hipótesis, la condensación por ParB es una consecuencia del atrapamiento de ADN debido a la dimerización conjunta de los dominios N-terminal (DNT) y DCT. En este escenario la ausencia del DNT en la truncación del DCT resulta en una inhibición de la formación de complejos de ParB, y en consecuencia de la fuerza impulsora de la condensación, sin afección de la unión a ADN.

Con el objetivo de evaluar estas hipótesis, hemos iniciado experimentos con versiones mutantes de ParB y el DCT en un sistema combinado de PM con microscopía de fluorescencia por reflexión total interna (MF). La combinación de PM con MF nos permite monitorizar la unión de la proteína al ADN evitando la condensación con la fuerza aplicada. Nuestros experimentos preliminares indican una decoración del DCT en moléculas de ADN con ParB unido, y por tanto se favorece la segunda hipótesis. Este escenario también está apoyado por la ausencia de condensación en los mutantes en el DNT (R80A). Es necesario continuar investigando para determinar la unidad mínima de condensación. Para cumplir este objetivo es conveniente diseñar protocolos experimentales que permitan observar la formación de bucles de ADN como resultado de la interacción de protomeros de ParB. El uso de moléculas con dos secuencias *parS* es la aproximación a priori más obvia. Sin embargo, la baja especificidad por la secuencia *parS* requiere de un ajuste fino de las condiciones experimentales. De nuevo, la combinación de PM y MF es una poderosa herramienta para investigar las uniones de ParB a las secuencias *parS* a nivel de molécula individual. La localización estricta de ParB alrededor de las secuencias *parS* observada *in vivo* contrasta con la baja especificidad observada en experimentos bioquímicos y de molécula individual. Por tanto, los experimentos mencionados son críticos para comprender esa aparente discrepancia.

Nuestro objetivo a corto plazo es explorar el papel de ParB en la carga del complejo SMC. Se ha descrito que la interacción con ParB unido a *parS* es un requerimiento para la unión de las proteínas SMC al ADN. Dado que las proteínas SMC son elementos claves de la condensación del cromosoma y de la organización topológica del mismo, su falta de unión a ADN en ausencia de ParB es una característica intrigante de su actividad. Una posible explicación es que la estructura condensada alrededor de la secuencia *parS* sea requerida para la unión de SMC. Hay una gran variedad de modelos propuestos para explicar los mecanismos de unión y condensación del ADN por parte de las proteínas SMC, y actualmente este es un tema muy activo y competitivo. Nuestros experimentos de PM son una valiosa herramienta para estudiar los mecanismos de la actividad de las proteínas SMC en presencia de ParB sobre las bases previamente adquiridas sobre ParB.

8.2 Dependencia de fuerza y torsion en la actividad de RepC

En la replicación por el mecanismo conocido como círculo rodante, las proteínas iniciadoras de la replicación están involucradas en la generación de un corte en hebra simple en el origen de replicación de los plásmidos. Esto induce la carga de la maquinaria de replicación y la síntesis de ADN comienza a partir del punto de corte. Los

detalles mecanísticos de este proceso no están bien conocidos a pesar de la implicación de de los plásmidos en la biología de Bacteria. En este proyecto hemos estudiado la actividad de corte de la proteína RepC en función del grado de superenrollamiento del ADN.

8.2.1 Conclusiones

Hemos diseñado un protocolo experimental para medir la dependencia del superenrollamiento en la actividad de RepC (Capítulo 6). Encontramos que la incubación de RepC con moléculas con superenrollamiento negativo resulta en la liberación de hasta 30 vueltas en un solo paso. Por el contrario, nunca encontramos actividad de corte en moléculas con superenrollamiento positivo. Hemos explotado esta característica pare medir precisamente el grado de superenrollamiento negativo necesario para iniciar la actividad de RepC. La proteína RepC se incubó con moléculas de ADN superenrolladas positivamente para promover la actividad de corte. Medimos el grado de superenrollamiento que daba lugar al corte por RepC, encontrando una dependencia con la fuerza aplicada. En cierta manera sorprendente, se observó que a fuerzas más altas se requieren menos vueltas para iniciar la actividad de RepC, al menos para el rango de fuerzas estudiadas en este trabajo (0.05 pN a 1 pN). El análisis de estos datos indica energías de superenrollamiento compatibles con las energias de activación descritas para la extrusión de estructuras cruciformes. La presencia de secuencias que potencialmente pueden resultar en estructuras cruciformes ha sido descrita previamente por otros grupos. Nuestros resultados apoyan un modelo donde la actividad de la proteína iniciadora de la replicación depende de la extrusión de una estructura cruciforme.

También investigamos la naturaleza de la interacción entre el ADN y RepC, a priori y a posteriori de la reacción de corte. Comparamos la actividad de corte de RepC con la de la topoisomerasa humana IB (Topo IB). Dicha enzima relaja el estrés torsional del ADN por un mecanismo análogo al de RepC. Sin embargo, encontramos que el comportamiento de esta enzima es muy distinto al de RepC. Mientras que Topo IB muestra una eficiente religación, raramente detectamos eventos de religación con RepC. Además, la actividad de la proteína Topo IB es independiente del grado de superenrollamiento en el ADN, contrariamente a nuestras observaciones para RepC. Esto indica que, mientras que ambas enzimas tienen mecanismos catalíticos similares, el mecanismo de interacción con el ADN debe ser diferente. Combinando nuestros experimentos de molécula individual con modelado por homología, concluimos que RepC es un dímero con una estructura en forma de C.

Proponemos un modelo mecanístico simplificado, donde el superenrollamiento desencadena la extrusión de una estructura cruciforme. La extrusión de la estructura cruciforme posiciona el punto de corte en el residuo catalítico de RepC. Nuestros resultados sugieren una extrusión espontanea de la estructura cruciforme bajo las condiciones adecuadas de fuerza y torsión, contrario a los modelos previos donde se RepC tiene un papel activo en la extrusión de la estructura.

Las conclusiones principales de este proyecto se resumen en los siguientes puntos:

- La actividad de RepC es dependiente del estado de superenrollamiento del ADN. Encontramos que es necesario superenrollar negativamente el ADN para desencadenar la actividad de RepC. Es interesante notar que, mientras que RepC es una eficiente enzima de corte en ADN negativamente superenrollado, nuestros resultados muestran una baja actividad de religación. Estos resultados contrastan con los mismos experimentos realizados con topoisomerasas, cuyo mecanismo de corte es similar al de RepC.

- La actividad de RepC es dependiente de la fuerza aplicada sobre el ADN. Nuestros resultados indican que para iniciar la actividad de RepC se requieren más vueltas a bajas fuerzas que a altas fuerzas. Las energías de superenrollamiento son consistentes con la formación de una estructura cruciforme.

- Los experimentos de PM así como el modelado de homología indican una estructura en forma de C para el dímero de RepC. Proponemos un modelo donde la extrusión de una estructura cruciforme posiciona el sitio de corte en el ADN cerca del residuo catalítico de RepC.

8.2.2 Perspectivas futuras

Nuestros resultados son consistentes con la formación de una estructura cruciforme como mediador de la actividad de RepC. Sin embargo, nunca hemos sido capaces de medir la formación de estas estructuras. Por tanto, no se puede descartar que otras estructuras transitorias o regiones de ADN desnaturalizado sean suficientes para la actividad de RepC. Para comprender las implicaciones de estructuras secundarias de ADN en la actividad de RepC, sería útil hacer experimentos en sustratos de ADN que no formen estructuras cruciformes, sin alterar el contenido de GC.

Además, el uso de moléculas más cortas sería conveniente para resolver cambios en extensión atribuibles a la extrusión de la estructura cruciforme o de otras estructuras secundarias alternativas. Ajustar las condiciones experimentales para resolver la estructura cruciforme sería una herramienta de gran utilidad para cuantificar las implicaciones de estas estructuras en la unión y la actividad de corte de RepC.

La incorporación del resto de componentes involucrados en la replicación es un paso adelante para el entendimiento de la replicación por círculo rodante. La siguiente enzima de la reacción después del corte de RepC, es la helicasa PcrA, cuya actividad es de hecho altamente dependiente de RepC. Dicha proteína carga PcrA en el origen de replicación y separa la doble hélice de ADN empezando desde el punto de corte de RepC. La producción de ADN de hebra simple (ADNhs) se puede monitorizar con pinzas magnéticas puesto que –a las fuerzas aplicadas en este trabajo– las diferentes propiedades mecánicas del ADNhs y ADN de hebra doble dan lugar a una reducción de extensión. En consecuencia es posible medir a nivel de molécula individual la actividad de la helicasa PcrA en complejo con RepC. La Dra. Carolina Carrasco

ha optimizado las condiciones experimentales para este estudio, obteniendo interesantes resultados preliminares. La caracterización de la actividad de PcrA así como sus implicaciones en la actividad de corte de RepC serán el objeto de una próxima publicación.

List of publications

- J.A. Taylor†, **C.L. Pastrana**†, A. Butterer, C. Pernstich , E.J. Gwynn, F. Sobbot, F. Moreno-Herrero , and M.S. Dillingham. Specific and non-specific interactions of ParB with DNA: implications for chromosome segregation. *Nucleic Acids Res.* 43(2):719-731 (2015).

- **C.L. Pastrana**†, C. Carrasco†, P. Akhtar, S.H. Leuba, S.A. Khan, and F. Moreno-Herrero. Force and twist dependence of RepC nicking activity on torsionally-constrained DNA molecules. *Nucleic Acids Res.* 44(18):8885-8896 (2016).

- G.L.M. Fisher†, **C.L. Pastrana**†, V.A. Higman, A. Koh, J.A. Taylor, A. Butterer, T.D. Craggs, F. Sobott, H. Murray, M.P. Crump, F. Moreno-Herrero and M.S. Dillingham. The C-terminal domain of ParB is critical for dynamic DNA binding and bridging interactions which condense the bacterial centromere. Under review at *eLife*.

- J. Madariaga-Marcos, S. Hormeno, **C.L. Pastrana**, G.L.M. Fisher, M.S. Dillingham, and F. Moreno-Herrero. Force determination in Lateral Magnetic Tweezers combined with TIRF microscopy. Under review at *Nanoscale*.

- Dynamics of nicking and unwinding of DNA by the RepC-PcrA complex. Carrasco C., **C.L. Pastrana**, S.H. Leuba, S.A. Khan, and Moreno-Herrero F. In preparation.

- Implications of the C-terminal domain of ParB in DNA condensation visualized by MT-TIRF under controlled force. J. Madariaga-Marcos, G.L.M. Fisher, **C.L. Pastrana**, M.S. Dillingham, and F. Moreno-Herrero. In preparation.

† These authors contributed equally to this work.

Part V

Appendixes

A

Force-extension relation in the Freely-jointed chain model

A.1 Introduction

In the section 2.4 (Chapter 2), different polymer models were presented. In particular, it was emphasized the relation between force and extension due to its practical application to single-molecule experiments. The Freely-jointed chain (FJC) is a relatively simple model that has been probed to accurately describe the force-extension relation for ssDNA [Bosco et al., 2014; Jarillo et al., 2017]. However, it does not fit the force-extension response for double-stranded DNA or proteins at *intermedium* forces (0.1 pN < F < 5 pN). Despite these limitations, the simple concept of the model allows to determine analytically –and without approximations– an expression for the extension as a function of the applied force [Grosberg and Khlokhlov, 1994]. In this appendix, a complete derivation of the force-extension relation presented in Equation 2.5 is shown.

A.2 Derivation

The Freely-jointed chain defined in section 2.4.1 considers the polymer as a set of independent segments of length b. The total energy of the chain U in the presence of an stretching force F is only dependent on the alignment of its segments with the force axis,

$$U = -Fb\sum_{i=1}^{N}\hat{\mathbf{t}}_i\cdot\hat{\mathbf{u}}_{\mathbf{F}} = -Fb\sum_{i=1}^{N}\cos\phi_i, \tag{A.1}$$

where $\hat{\mathbf{t}}_i$ is the unitary orientation vector, $\hat{\mathbf{u}}_{\mathbf{F}}$ states for the unitary vector in the direction of the applied force, and ϕ_i is the angle between the ith segment and the stretching force axis (Figure A.1). That is to say, the bending energy is null and in the presence of an external force, there is an energetic penalty due to the alignment of the segments with $\hat{\mathbf{u}}_{\mathbf{F}}$ against the entropic force of the chain. For simplicity, the force axis is assumed to be z axis. Then, $\hat{\mathbf{u}}_{\mathbf{F}} = \hat{\mathbf{k}}$ axis and ϕ is a polar angle.

The partition function $\mathcal{Z}$ of the system considering all the possible configurations can be written as

$$\mathcal{Z} = \sum_{\Omega} e^{\frac{Fb}{k_BT}\sum_{i=1}^{N}\cos\phi_i} = \sum_{\Omega}\prod_{i=1}^{N} e^{\frac{Fb}{k_BT}\cos\phi_i}, \tag{A.2}$$

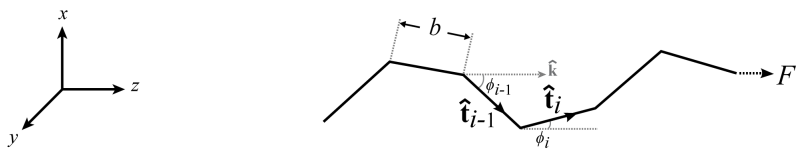

Figure A.1: Freely jointed-chain model. The polymer is described as a set of segments of size b with orientation vector $\hat{\mathbf{t}}_i$ forming an angle ϕ_i with the axis of the applied force.

where the summatory over Ω indicates the full conformational space, whilst the productory involves all the segments of a given chain [Strick et al., 2000]. Since the segments are independent, equation A.2 can be defined as

$$\mathcal{Z} = \left[\sum_{\Omega} e^{\frac{Fb}{k_B T} \cos\phi} \right]^N . \tag{A.3}$$

For a single segment, the full conformational space can be written in spherical coordinates [Cantor and Schiemel, 1980]. The partition function of this segment $\mathcal{Z}_s$ such that $\mathcal{Z}_s^N = \mathcal{Z}$ is then given by

$$\mathcal{Z}_s = \int_0^{2\pi} d\theta \int_0^{\pi} e^{\frac{Fb}{k_B T} \cos\phi} \sin\phi d\phi, \tag{A.4}$$

where θ and ϕ are the azimuthal and polar angles respectively. This double integral can be solved analytically by substitution considering $u = Fb(k_BT)^{-1}\cos\phi$ and $du = -Fb(k_BT)^{-1}\sin\phi$,

$$\begin{aligned} \mathcal{Z}_s &= -2\pi \frac{k_B T}{Fb} \int e^u du \\ &= -2\pi \frac{k_B T}{Fb} e^{\frac{Fb}{k_B T}\cos\phi} \Big|_{\phi=0}^{\phi=\pi} \\ &= 2\pi \frac{k_B T}{Fb} (e^{Fb/k_BT} - e^{-Fb/k_BT}). \end{aligned} \tag{A.5}$$

The last expression can be written in a more compact form as,

$$\mathcal{Z}_s = 4\pi \frac{k_B T}{Fb} \sinh\left(\frac{Fb}{k_B T}\right), \tag{A.6}$$

and then,

$$\mathcal{Z} = \left[4\pi \frac{k_B T}{Fb} \sinh\left(\frac{Fb}{k_B T}\right) \right]^N \tag{A.7}$$

The average extension of the molecule l can be determined from the relation

$$l = -\frac{\partial H}{\partial F}, \tag{A.8}$$

where H is the free energy of the polymer given by $H = -k_B T \log \mathcal{Z}$. This is,

$$\begin{aligned} l &= k_B T \frac{\partial \log \mathcal{Z}_s^N}{\partial F} \\ &= k_B T N \frac{\partial \log \mathcal{Z}_s}{\partial F}. \end{aligned} \tag{A.9}$$

The partial derivative can be calculated as

$$\begin{aligned} l &= k_B T N \frac{d \log \mathcal{Z}_s}{d \mathcal{Z}_s} \frac{\partial \mathcal{Z}_s}{\partial F} \\ &= k_B T N \frac{Fb}{4\pi k_B T \sinh\left(\frac{Fb}{k_B T}\right)} \left[\frac{4\pi}{F} \cosh\left(\frac{Fb}{k_B T}\right) - 4\pi \frac{k_B T}{F^2 b} \sinh\left(\frac{Fb}{k_B T}\right) \right] \end{aligned} \tag{A.10}$$

that can be simplified, yielding

$$l = Nb \left[\coth\left(\frac{Fb}{k_B T}\right) - \frac{k_B T}{Fb} \right]. \tag{A.11}$$

Since Nb equals the total length of the chain ($L = Nb$), the derivation of the result shown in expression 2.5 is finished. In this model, b is usually termed *Kuhn length*, and equals two times the persistence length of the polymer.

The assumption of a negligible bending energy is not valid for DNA, and therefore this model poorly describes the experimental data for DNA. The FJC model only provides satisfactory results of the force-extension response for polymers with small bending modulus, such that the stretching energy strongly dominates over the bending energy. However, the simplicity of the FJC provides a useful and didactic framework to tackle and derive an analytical solution of a polymer model.

A.3 References

Bosco, A., Camunas-Soles, J., and Ritort, F. Elastic properties and secondary structure formation of single-stranded DNA at monovalent and divalent salt conditions. *Nucleic Acids. Res.*, 42(3): 2064-74, 2014.

Cantor, C. R., and Schimmel, P. R. *Biophysical Chemistry, Part III: The Behaviour of Biological Macromolecules.* W.H. Freemann, San Francisco, 1980.

Grosberg, A. Y., and Khlokhlov A. R. *Statistical Physics of Macromolecules.* AIP, New York, 1994.

Jarillo, J., Morín, J. A., Beltrán-Heredia, E., Villaluenga, J. P. G., Ibarra, B. and Cao, F. C.. Mechanics, thermodynamics, and kinetics of ligand binding to biopolymers. *PLoS One*, 12(4):e0174830, 2017.

Strick, T., and Allemand, J., and Croquette, V., and Bensimon, D. Twisting and stretching single DNA molecules. *Prog. Biophys. Mol. Biol.*, 74(1-2): 115-40, 2000.

B

Allan variance analysis of MT spatial and temporal resolution

B.1 Introduction

The applied force in magnetic tweezers is determined from calibration mesurements. These are dependent on accurately sampling the transversal fluctuations of the bead and the length of the tether (see Chapter 3). It is therefore necessary to measure for sufficiently long time to ensure non-correlated measurements and proper estimation of the coordinates of interest and derived parameters. However, thermal and mechanical drifts during experiments have a contribution to the measured signal in the form of low frequency noise, and thus data acquisition for very long times is not a recommended solution. In order to find an optimal measurement time such that instrumental noise and accurate sampling of beads positions are balanced, a convenient and simple approach relies in the analysis of the Allan variance (AV) [Czerwinski et al., 2009]. The (AV) $\sigma_x^2(\tau)$ for a set of N data points binned in intervals of k points, is defined as,

$$\sigma_x^2(\tau) = \frac{1}{2}\left\langle (\bar{x}_n - \bar{x}_{n+1})^2 \right\rangle_\tau , \tag{B.1}$$

where τ is the measurement interval window ($\tau = f_{acq}k$), $\bar{x}$ is the average of the interval, and the brackets $\langle ... \rangle$ indicate the average of the difference for the N/k intervals. That is to say, the AV is half the average of the square difference of adjacent intervals of time length τ. The AV can be calculated considering consecutive bins sharing data points, and in this case it is said to be overlapping [Lansdorp and Saleh, 2012]. From the analysis the AV of variable τ the acquisition time minimizing the variance of the data and its associated spatial resolution can be resolved.

B.2 Results

For the analysis of the AV we obtained data sets at 400 Hz for a total time of 500 seconds on ≈2μm contour length molecules. Notice that although the AV is dependent on the acquisition frequency, minor differences are expected for the range of available frequencies in our MT setup [Czerwinski et al., 2009]. The results for the analysis of the non-overlapping AV for at least 4 beads at 23 different time windows τ from 0.005 seconds (k=2 data points) to 200 seconds (k=8·10^4 points) are shown in Figure B.1.

For high forces ($F = 12$ pN), a maximum spatial resolution of ≈2 nm is obtained

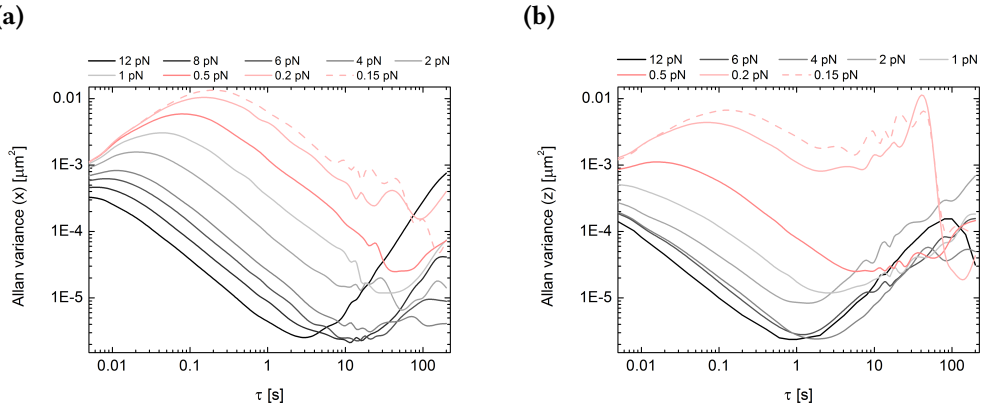

Figure B.1: Spatial resolution and temporal stability characterized by the Allan variance. a) Allan variance at the specified forces for the x axis. b) Allan variance at the specified forces for the z axis. The data represent the average of least 4 molecules. 500 seconds data sets were acquired at 400 Hz in molecules of 2 μm contour length.

for measurements of ∼1-2 seconds long for both the xy axes (Figure B.1a) and the z axis (Figure B.1b). Similar results are obtained for surface-stuck beads, indicating the resolution limit of the setup.

The reduction of the force results in an increase of the sampling time, reaching tens to hundreds of seconds for $F < 0.5$ pN. Interestingly, the variance for the x and z axis for forces above 4 pN remains unchanged. This result can be explained by the reduction of the bending fluctuations, since at 4 pN the extension of the molecule is ∼85% of its total contour length.

The measurements of the z axis at low forces $F < 0.5$ pN are more senstive to instrumental noise than for the x axis, as reflected by the lost of tendency for $\tau > 2$ s. An accurate characterization of the AV for the z axis at this low forces requires acquisition of longer data sets to increase the number of available time intervals for analysis. Nonetheless, the results suggest requering up to hundred of seconds.

In summary, the AV provides a powerful tool to analyze both the resolution and the stability of the setup, as well as to select an optimum measuring time minimizing the instrumental drift without compromising an accurate sampling.

B.3 References

Czerwinski, F., Richardson, A. C., and Oddershede, L. B. Quantifying noise in optical tweezers by Allan variance. *Opt. Express*, 17(15):13255-69, 2009.

Lansdorp, B. M., and Saleh, O. A. Power spectrum and Allan variance methods for calibrating single-molecule video-tracking instruments. *Rev. Sci. Instrum*, 83(2): 025115, 2012.

C

A simple model to characterize drag and DNA extension in the presence of flow and magnetic forces

C.1 Introduction

Several magnetic tweezers (MT) experiments described in the results section were conducted under the action of a flow and a stretching magnetic force. This appendix describeS a simplified model attending to the geometrical constrains of the DNA-bead system to characterize the force of an applied flow. The model is used to characterize the total force acting on the DNA and its effect on the measured extension as a result of combined stretching and drag forces.

C.2 Model

The forces acting on a DNA molecule tethered to a bead subjected to a magnetic stretching force on the z axis, ($\mathbf{F_m}$) and a drag force on the x axis due to the liquid flow ($\mathbf{F_d}$) are reppresented in Figure C.1a. The drag force acting on the bead ($\mathbf{F_d} = F_d\hat{\mathbf{i}}$) as a result of the applied flow is given by Stokes' law:

$$F_d = 6\pi\eta_c R v \tag{C.1}$$

where η_c is the viscous coefficient of the buffer corrected for the distance to the surface, R is the diameter of the bead and v is the velocity of the fluid. Therefore, the resulting force vector $\mathbf{F_t}$ acting on the bead is given by $\mathbf{F_t} = 6\pi\eta R v\hat{\mathbf{i}} + F_m\hat{\mathbf{k}}$. In order to determine the drag force it is first necessary to calculate the viscous coefficient η_c. The viscosity near surfaces can be significantly larger than the bulk viscosity η when the distance between the surface and the bead, z, is comparable to R. This situation is common in MT experiments, where the molecular extension of the DNA l_0 is a few micrometers. Then, for the bead-DNA system, $z = l_0 + R$. The viscosity at a distance z to the surface can be found considering Faxen's corrections [Happel and Brenner, 1983]:

$$\eta_c(z) = \frac{\eta}{1 - \frac{9}{16}\left(\frac{z}{R}\right) + \frac{1}{8}\left(\frac{z}{R}\right)^3 - \frac{45}{256}\left(\frac{z}{R}\right)^4 - \frac{1}{16}\left(\frac{z}{R}\right)^5} \tag{C.2}$$

The extension of the molecule l_0 in absence of drag ($F_d = 0$) can be found from the Moroz and Nelson interpolation formula for the worm-like chain (WLC) described

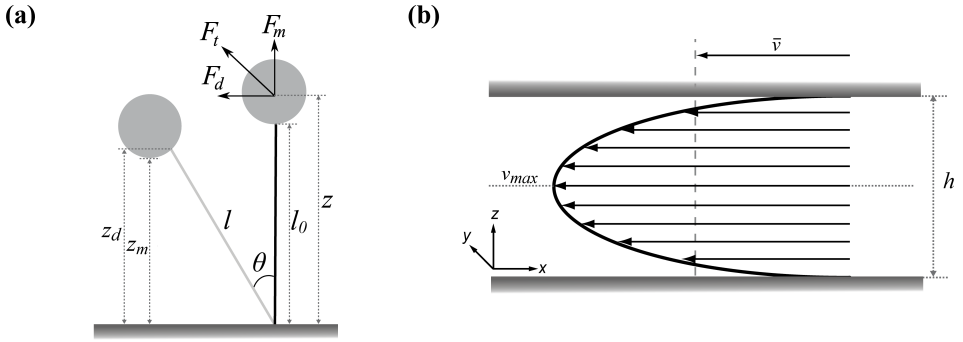

Figure C.1: Combined stretching and drag forces in MT experiments. a) Diagram of forces acting on the bead and related quantities. b) Laminar flow profile of velocities for a circular pipe. Similar profiles are obtained in rectangular MT microfluidics cells.

in section 2.4.2, including an elastic component:

$$l_0(F_m) = L\left[1 - \frac{1}{2}\left(\frac{F_m\xi}{k_BT} + \frac{1}{32}\right)^{-1/2} + \frac{F_m}{S}\right], \tag{C.3}$$

where ξ is the persistence lenght of the DNA and S is the stretching modulus.

In addition, to extract the exerted force by the flow drag, it is necessary to determine the velocity inside the liquid chamber from the flow profile as a function of the distance to the surface $v(z)$ [Brewer and Bianco, 2008]. The control parameter in the syringe pump used for the exchange of buffers is the flow rate, Q. Then, the average velocity inside the liquid chamber is $\bar{v} = Q/A$, where A is the cross-sectional area. For our liquid chambers the width, $w = 7$ mm and the height, h, of a single-parafilm layer chamber is 0.125 mm or 0.250 mm for a double parafilm layer. Although the theory of laminar flows is well defined for circular pipes, it is necessary to apply empirical approximations for the determination of $v(z)$ in other geometries [Çengel and Cimbala, 2013]. This is the case of the rectangular duct of the microfluidics liquid chamber employed in our MT setup. The assumption of laminar flow can be justified attending to the Reynolds number (Re) given by

$$Re = \frac{2\rho\bar{v}r_h}{\eta} \tag{C.4}$$

where ρ is the density of the fluid, η is the viscosity and r_h is the hydraulic radius. For a rectangular pipe $r_h = hw/(h + w)$, obtaining $Re \approx 9.5$ for the maximum flow rate of our syringe pump ($Q = 999$ µl/min). Since $Re \ll 2000$, we can assume laminar flow conditions.

The flow velocity profile through the center of a rectangular pipe ($y = w/2$) and at a certain distance to the surface z can be obtained following the approximations described in Shah [1978] as

$$v(z) = \bar{v}\frac{3}{2}\frac{m+1}{m}\left[1 - \left(\frac{h-2z}{h}\right)^2\right]. \tag{C.5}$$

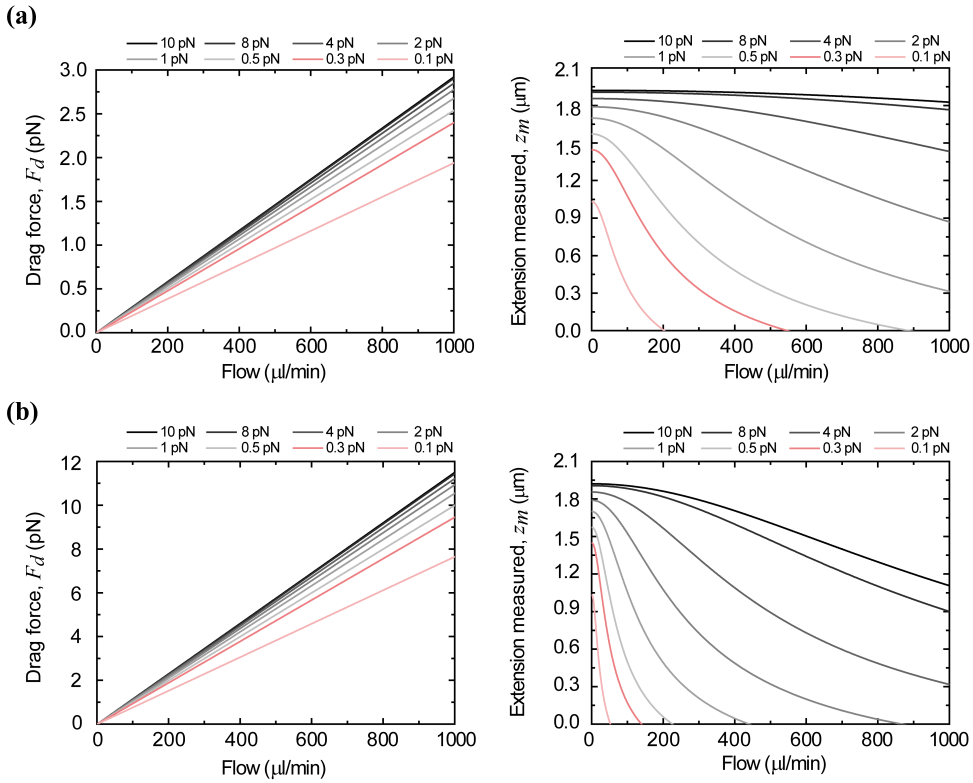

Figure C.2: Characterization of combined magnetic and drag stretching forces on DNA molecules. Results for the drag force (F_d) and the measured extension in the MT setup (z_m) for double-parafilm layer cells (a) and for single parafilm layer cells (b). The increase of the drag force with the stretching force is the result of the increase of the velocity with the distance to the surface. Parameters used: Contour length $L = 2$ μm, persistence length $\xi = 45$ nm, bead radius $R = 500$ nm, stretch modulus $S = 1200$ pN, and bulk viscosity $\eta = 1$ mPa·s.

In the previous expresion m is an empirically-determined constant given by

$$m = 1.7 + \frac{1}{2}\alpha^{-1.4} \tag{C.6}$$

being $\alpha = h/w$ the aspect ratio of the chamber. The design of the microfluidic cells used in MT experiments –where $h \ll w$– results in parabolic profiles similar to that described for circular pipes (Figure C.1b). Finally, it is possible to determine the drag force as $F_d(z(F), v(z)) = 6\pi\eta_c(z)Rv(z)$.

The sum of the magnetic force and the drag force results in a total force $\mathbf{F_t}$ which stretches the molecule with an angle θ with respect to vertical axis. The stretching of the DNA by $\mathbf{F_t}$ increases the end-to-end distance of the molecule $l(F_t)$, and this can be found from the WLC model (equation C.3) substituting the magnetic force, F_m, by $|\mathbf{F_t}| = \sqrt{F_m^2 + F_d^2}$.

The angle with the vertical axis θ is obtained from $\cos\theta = F_m|\mathbf{F_t}|^{-1}$ and the z axis component of the extension is determined as $z_d = l\cos\theta$. Nevertheless, the

pendulation of the bead results in an understimation of the measured extension of the bead due the *hidden* extension by the radius of the bead (see sections 3.4.2 and 3.7.2). Considering south-pole attachment of the DNA, the measured extension in the MT software is obtained by subtracting the contribution of the bead radius as $z_m = z_d - R(1 - \cos\theta)$. For simplicity, this analysis considers DNA attachment to the south pole of the bead.

The solution to the previous equations for a characteristic 2 μm DNA molecules used in MT experiments is represented in Figure C.2. As expected, the lower velocities for a given flow in cells with two parafilm layer (Figure C.2a) compared with cells made with one parafilm layer (Figure C.2b) results in lower drag forces. Importantly, the non-linear dependency of the velocity with the height of the cell leads to much larger forces in cells fabricated with a single-parafilm layer than when two parafilm layers are used for its assembly. Consequently, shorter extensions are measured for the same flow in single-parafilm layer cells. Moreover, the non-linear mechanical response of the DNA results in a more dramatic reduction of the extension with the applied flow at low magnetic forces than at higher forces ($F > 1$ pN), where bending fluctuations are suppressed by the magnetic stretching force.

Remarkably, there is a plateau region where the flow barely affects the extension measured z_m. For instance, at 1 pN the flow in double-parafilm layer cells can go up to ∼100 μl/min before the extension l is significantly affected by the drag force (Figure C.2a). It is important to note that therefore, the lack of reduction of extension cannot be used as a hallmark of a negligible drag compared with the stretching force. For instance, at a stretching force of 1 pN, and using double-parafilm layer cells, the extension is reduced only a 4% for a ∼100 μl/min flow, but the drag force equals 0.26 pN. Special care must be taken to avoid exert drag forces comparable (or higher) than the magnetic force, since this can affect the determination of the mechanical properties of the molecule or prevent the binding of proteins. The experiments described in this thesis used a flow rate of 18 μl/min at most. This results in drag forces of 0.04 pN for DNA molecules stretched with a magnetic force of 0.35 pN, or to a drag forces of ∼ 0.02 pN in supercoiled or condensed molecules (for the same magnetic force).

In summary, the method here described can be used as a reference of the applied force on the DNA during buffer exchange and as its consequences for the measured coordinates during data acquisition.

C.3 References

Brewer, L. R. and Bianco P. R. Laminar flow cells for single-molecule studies of DNA-protein interactions. *Nat Methods*, 5(6):517-25, 2008.

Çengel, Y. A. and Cimbala, J. M. *Fluid mechanics: fundamentals and applications*. McGraw-Hill, 3rd edition, 2013.

Happel, J. and Brenner, H. *Low Reynolds number hydrodynamics*. Martinus Nijhoff Publishers. 2nd edition, 1983.

Shah, R. K., London, A. L., Irvine, T. F., and Hartnett, J. P. *Laminar flow forced convection in ducts. A source book for compact heat exchanger analytical data.* Academic Press, 1st edition, 1978.

D

Implementation of C/C++ libraries to run under LabVIEW

D.1 Introduction

LabVIEW is a widely-used programming language in the scientific community for machine control and data analysis. Its intuitive environment combines an easy programming with a large variety of built-in functions and a manifold of tools for creating graphical user interfaces (GUI). However, it is restricted by its slowness in processing large amounts of data or to carry complex mathematical calculations and this is intrinsic to its interpreted language condition.

In contrast, compiled programming languages, e.g. C or Fortran, are the best option for simulations and numerical evaluation. Unfortunately, these programming languages involve a deeper understanding of the internal architecture of the computer and the creation of GUI is neither fast nor straightforward.

The inclusion of compiled libraries in LabVIEW is a beneficial solution to combine the power of both approximations. This appendix illustrates the basic guidelines for the creation of a C-written dynamic library (DLL in the Windows systems) and its calling under the LabVIEW environment. For didactic purposes, the explanation takes as example the generation of a step fitting routine. In the example, the LabVIEW environment passes a two-dimensional array containing the data to fit to the C library, and this returns to LabVIEW the result of the fitting in the form of a two dimensions array. The objective is to show a complex situation that can be used as a reference for the rational to follow for other purposes.

D.2 Methods

For the step fitting algorithm, a running-window of the data is fitted to a function of the form $z(t) = z_0 + S \cdot \mathcal{H}(t - t_s)$, where S is the step size and $\mathcal{H}$ is the Heaviside step function, which returns 0 for $t < t_s$ and 1 otherwise. The best fit is obtained by trying a range of possible values provided by the user. The data to fit corresponding to experimental data is previously selected and analyzed with a LabVIEW software. After calling the C library for the fit, the data is returned to LabVIEW for the plotting and subsequent treatment of the data.

C side

Many integrated development environments (IDE) incorporate the option to create a library project and automatically configure the compiler for that purpose. We have used Dev-C++ with the MinGW compiler, the GCC implementation for the Windows operative system. This compiler requires including the `windows.h` header, and declaring the functions that will be accessible from outside as exportable with the sentence `__declspec (dllexport)`. The inclusion of functions for calling by the Windows' API is also necessary as well as the parameters for compiling to a dynamic library. Since this is dependent on the compiler, the details are skipped. Look at the documentation of your compiler for a detailed description of the procedure.

The prototype of the main function of the library for the fitting algorithm is defined as

```
__declspec (dllexport) int *heaviside_fit( double *input_data, int points,
    double S_MAX, double S_MIN, double RES_STEP, double MAX_OFFSET, double
    MIN_OFFSET, double RES_OFFSET, double CHI_SQRT, int window_size );
```

The function takes as input arguments a vectorized two-dimensional array containing the time and extension from magnetic tweezers data `*input_data` and the number of points `points`. The arguments `S_MAX`, `S_MIN`, `RES_S` are the minimum and maximum step S possible and the resolution of the step. Similarly, `MIN_OFFSET`, `MAX_OFFSET`, and `RES_OFFSET` are the minimum and maximum offset z_0 and the resolution in the determination of the offset. `CHI_SQRT` states for a tolerance criterion of the fit, and `window_size` defines the number of points in the running window.

Notice that while `*input_data` has two dimensions, it has been defined as a one-dimension array. Since C is a row-major language, the access to the two-dimension matrix in a vectorized form is possible by considering that the rows of the matrix are contiguous in memory (Figure D.1). In other words, to access the element in row `i` and column `j`, we have to write `A[i*N + j]`, where `N` is the number of columns. This expresion can be deduced from the structure presented previously for two dimensional arrays. Notice that these concepts are especially important to understand the procedures followed in the next steps.

The function returns a two-dimensional array containing the time, the extension as a result of the step fitting procedure, and the parameters S and z_0 at each t. The function to be returned is first allocated dynamically by calling the `malloc()` function as

```
double **fitted_steps = (double **)malloc(POINTS * sizeof(double) );
for (int i = 0; i < points; i++)
  fitted_steps[i] = (double *) malloc( 4 * sizeof(double*) );
```

In summary, a matrix of `points` rows and `4` columns is allocated as a pointer to an array of pointers. After the fitting routine, the data of the analysis is returned simply as `return fitted_steps`. Hence, the value returned is the location in memory of the first element of the matrix.

(a)

1,1	1,2	1,3
2,1	2,2	2,3
3,1	3,2	3,3

(b)

1,1	1,2	1,3	2,1	2,2	2,3	3,1	3,2	3,3

Figure D.1: Row-major order of matrix store. The 3x3 matrix in panel (a) is stored as contiguous elements in memory as represented in panel (b). Alternatively, languages with column-major order (e.g. Fortran) the elements in columns are contiguous in memory.

LabVIEW side

LabVIEW 2011 was used for the development of the procedure exposed here. The calling of DLLs in LabVIEW is performed through the `Call Library Function Node VI` (Figure D.2). Once the VI has been included it is necessary to configure the parameters by double-clicking on the control. In the opened dialog, go to the tab Function. Select the localization of the DLL and the name of function that is going to be used, in this case `heaviside_fit`. In case of requiring using different functions of the same DLL, it will be necessary to include additional `Call Library Function Node` functions. Importantly, if the software is shared among different users, the path needs to be updated for each computer, since LabVIEW does not consider relative paths for the DLL. In the same dialog configure the parameters of the function for all the inputs in the Parameters tab selecting the data type of each function of the library. LabVIEW can communicate directly with C for single values of **`int`**, **`float`** **`double`**, and **`char`**. The description of how to input a two-dimmensional array is shown in Figure D.3a. In Figure D.3b, it is represented the wiring of the Call Library Function Node VI for the example described. The input matrix corresponds to low-pass filtered data selected by the user on a plot with cursors. The remaining variables are introduced with numeric control boxes. The output parameter `pointer_LVHandle` contains the memory address of the first element of the `fitted_steps` array returned in the C library. As a consequence, it is necessary to dereference this pointer in order to extract the values for all the positions of the array.

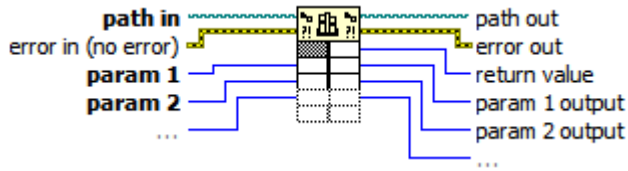

Figure D.2: `Call Library Function Node` VI input and output. It is necessary to add as many racks as inputs has the function in the C library. This results in a VI with a number of outputs equal to the number of inputs plus one. **`return`** `value` is the value returned by the C function. The `param n output` returns the value at the position of the memory of `param n`. This allows returning more than a single variable from the library.

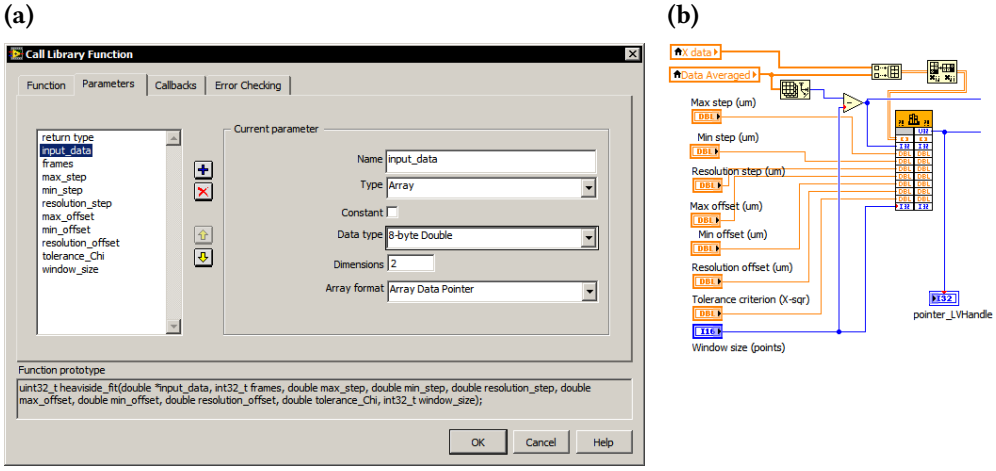

Figure D.3: Configuring the `Call Library Function Node` VI. a) Parameters for inputting a two-dimensional array. b) The aspect of the VI once configured and wired for the example.

LabVIEW provides different approaches to dereference pointers. The first option, is to directly configure the `Call Library Function Node` VI for this purpose. If the variable of interest is not the returned one, the value of the pointer can be read by changing the configuration in the Parameters tab, selecting `Signed Pointer-sized` Integer in the Data type box and `Pointer to Value` in the Pass combo box.

In the case of the returned variable, there are two alternatives. A first approach is to use an external LabVIEW DLL, termed `MoveBlock` which have to be loaded and configured with the `Call Library Function Node` VI. The second option is to use the `GetValuebyPointer` VI. We find this alternative provides a more clear understanding of the process to dereference a two-dimensional array. Thus, here we are going to review only the methodology to dereference multidimensional arrays using the `GetValuebyPointer` VI. Both methods are described at the National Instruments online development center.

The `GetValuebyPointer VI` is analogous to the dereference or indirection operator `*` in C, i.e. given a memory address, it returns the value store at that block of memory (Figure D.4). However, in the present example, the memory address `pointer_LVHandle` corresponds only to the first element of the array (`fitted_steps [0][0]`). Two steps are necessary to fully dereference the array. Take present the structure in memory of the two-dimensional array described previously as well as the allocation procedure. The two-dimensional array has been created as pointer to an array of pointers.

In a first step, the dereferencing process involves a loop to obtain the memory address of the first element in each row of the matrix (Figure D.5, *Dereference 1*). Thus, the result of the first dereferencing step involves obtaining an array of memory addresses. For this, a N iterations loop is iterated to dereference the memory

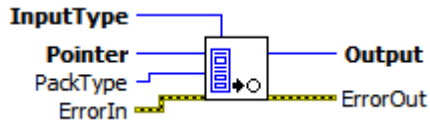

Figure D.4: `Get Value by Pointer` VI. The VI takes as inputs a memory address `Pointer`, and the data type of the variable `InputType`, which is wired as a constant of the given data type. The VI returns (`Output`) the value at the specified memory address.

address m at $m = m_0 + i \cdot s$, where m_0 is the memory address of the first element (`pointer_LVHandle`), i is the iteration number, and s is the size in bytes of a pointer. For our 32-bits system, pointers are integers of 4 bytes. In summary, contiguous blocks of 32 bits are dereferenced to obtain a set of N memory addresses.

The next step involves dereferencing the columns of each row: every element of the array obtained previously containing the memory address of the first column is dereferenced for the subsequent columns (Figure D.5, *Dereference 2*). To extract the values of every column starting from the first element of the row, the memory is explored considering that the values of the columns are consecutive in memory. Therefore, starting from the memory address of the first column, the value of the next column is located at a number of bits equal to the size in bits of the data type. In the example described, `fitted_steps` is a **`double`** array. For a 32 bits system, a **`double`** variable is 64 bits (8 bytes) in size. Finally, the data produced externally in a C library has been incorporated in a two-dimensional array readable by LabVIEW.

Importantly, C arrays can be defined in different manners. The procedure described here is derived on the generation of pointer to an array of pointers. However, there are other allocation options. For instance,

```
double *fitted_steps = (double *)malloc(POINTS * COLUMNS * sizeof(double) );
```

allocates memory for one-dimension array, where the value of the element at row `i` and column `j` is obtained as `fitted_steps[i*COLUMNS + j]`. This procedure only requires a dereferencing loop in LabVIEW. However, the structure obtained is a vector and thus is necessary to post-process the data in LabVIEW to arrange a two-dimensional array.

In summary, there is not a unique solution, and the selection of a particular method depends on the preferred syntax by the programmer.

D.3 Conclusions

Here we have shown the basis to implement a dynamic library written in C and its communication with LabVIEW using an example with a didactic purpose. The details of the implementation of the C algorithm have been omitted to avoid that unnecessary details confuse the reader. For the same reason, the specific details for the compilation of dynamic libraries have also been omitted, to focus on the logic behind the communication between programming languages.

The communication between two different programming languages is a pow-

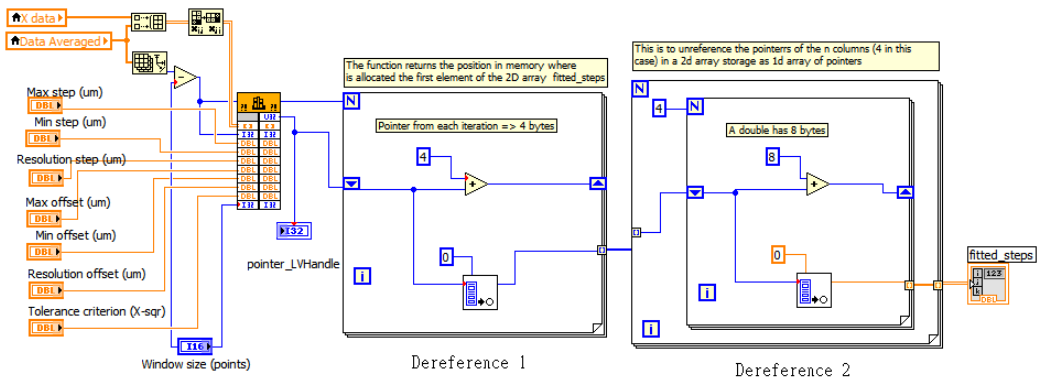

Figure D.5: Full diagram for calling a library and dereferencing a two dimensional array. The returned pointer is later dereferenced to an array of memory addresses of the first column for every row (*Dereference 1*). Next, from these memory addresses, the columns are dereferenced (*Dereference 2*) to obtain a two dimensions array.

erful tool to reuse codes for different tasks and combining different programming paradigms, obtaining the best features of each language. For the particular case of LabVIEW and C, this permits combining the computational power of C/C++ with the variety of graphical resources available for creating user interfaces in LabVIEW. Particularly, LabVIEW is widely used for the control of scientific instruments. Then, while the majority of the control can be performed directly by LabVIEW, the more computationally-demanding tasks can be derived to a C-written library.

Nonetheless, some considerations have to be studied before deciding to implement a C library. First, it is necessary to evaluate the time response of the communication between the library and LabVIEW to avoid efficiency bottlenecks at this stage. Second, care must be taken to ensure portability. Both LabVIEW and the library have to work with the same architecture, not being possible to have a coexistence of a 32 bits LabVIEW version with a 64 bits library. Despite nowadays the number of computers operating with a 32 bit architecture is reduced, is not uncommon to preserve obsolete systems in scientific equipments to avoid malfunction in case of updating. Therefore, the developer must consider this when selecting the compiler parameters. This appendix has described the procedures for implementing 32 bits libraries. The major difference for 64 bits library relies on the dereferencing process: the size of a `int *` variable is 8 bytes in size for 64 bits machines. Thus, the dereferencing process has to account for the size of the pointer. Alternatively, the size in bytes can be an additional parameter returned by the C library after evaluation of the `sizeof(int *)` expression.

In summary, coexistence of C and LabVIEW is not an immediate or simple procedure, requiring a notable effort to integrate both interfaces. However, is highly convenient for the control of sophisticated scientific equipment dependent on complex calculations, without sacrificing the advantages of LabVIEW to present and analyze data.

D.4 References

Ceballos, F. J. *Enciclopedia del Lenguaje C++*. Ra-Ma, 2005.

Goedecker, S. and Hoisie, A. *Performance Optimization of Numerically Intensive Codes*. Society for Industrial and Applied Mathematics, 2001.

E

List of abbreviations

AFM	.....	Atomic force microscopy
a.k.a.	.....	Also known as
anti-DIG	.....	Anti-digoxigenin IgG antibodies
API	.....	Aplication program interface
AV	.....	Allan variance
bp	.....	Base pairs
BSA	.....	Bovine serum albumine
CDBD	.....	Central DNA binding domain
CMOS	.....	Complementary metalŰoxideŰsemiconductor
CTD	.....	Carboxy terminal domain
DIG	.....	Digoxigenin
DNA	.....	Deoxyribonucleic acid
dsDNA	.....	Double-stranded DNA
DLL	.....	Dynamic-link library
DTT	.....	Dithiothreitol
sso	.....	Double-stranded origin of replication
EDTA	.....	Ethylenediaminetetraacetic acid
e.g.	.....	From latin, *exempli gratia* (for example)
EMSA	.....	Electrophoretic movility shift assay
FEC	.....	Force-extension curve
FFT	.....	Fast Fourier transform
FJC	.....	Freely-jointed chain
FOMT	.....	Freely orbiting magnetic tweezers
fps	.....	Frames per second
FRET	.....	Fluorescence resonance energy transfer
GCC	.....	GNU compiler collection
GUI	.....	Graphical user interface
Hepes	.....	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
IDE	.....	Integrated development environments
i.e.	.....	From latin, abbreviation of *id est* (it is)
KP	.....	Kratky-Porod model
LED	.....	Light-emitting diode
LUT	.....	Lookup table

MT	.....	Magnetic tweezers
nsDNA	.....	Non-specific DNA
NTD	.....	Amino (N) terminal domain
OFP	.....	Objective's focal plane
ori	.....	Origin of replication
OT	.....	Optical tweezers
PALM	.....	Photo-activated localization microscopy
PCR	.....	Polymerase chain reaction
PEEK	.....	Polyether ether ketone
RCR	.....	Rolling-Circle Replication
RNA	.....	Ribonucleic acid
ROI	.....	Region of interest
SC	.....	Supercoiled
s.e.m	.....	Standard error of the mean
SSB	.....	Single-strand DNA binding protein
ssDNA	.....	Single-stranded DNA
sso	.....	Single-stranded origin of replication
std	.....	Standard deviation
STED	.....	Stimulated emission depletion
STORM	.....	Stochastic optical reconstruction microscopy
TC	.....	Torsionally-constrained
TIR	.....	Total internal reflection
TPM	.....	Tether particle motion
Tris	.....	2-Amino-2-(hydroxymethyl)propane-1,3-diol
Tween	.....	Polyoxyethylene
WLC	.....	Worm-like chain